

Similarity Constraints-Based Structured Output Regression Machine: An Approach to Image Super-Resolution

Cheng Deng, Jie Xu, Kaibing Zhang, Dacheng Tao, *Fellow, IEEE*,
Xinbo Gao, *Senior Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

Abstract—For regression-based single-image super-resolution (SR) problem, the key is to establish a mapping relation between high-resolution (HR) and low-resolution (LR) image patches for obtaining a visually pleasing quality image. Most existing approaches typically solve it by dividing the model into several single-output regression problems, which obviously ignores the circumstance that a pixel within an HR patch affects other spatially adjacent pixels during the training process, and thus tends to generate serious ringing artifacts in resultant HR image as well as increase computational burden. To alleviate these problems, we propose to use structured output regression machine (SORM) to simultaneously model the inherent spatial relations between the HR and LR patches, which is propitious to preserve sharp edges. In addition, to further improve the quality of reconstructed HR images, a nonlocal (NL) self-similarity prior in natural images is introduced to formulate as a regularization term to further enhance the SORM-based SR results. To offer a computation-effective SORM method, we use a relative small nonsupport vector samples to establish the accurate regression model and an accelerating algorithm for NL self-similarity calculation. Extensive SR experiments on various images indicate that the proposed method can achieve more promising performance than the other state-of-the-art SR methods in terms of both visual quality and computational cost.

Manuscript received August 7, 2014; revised August 9, 2015; accepted August 10, 2015. Date of publication September 9, 2015; date of current version November 15, 2016. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB316400, in part by the National Natural Science Foundation of China under Grant 61125204, Grant 61572388, Grant 61432014, and Grant 61471161, in part by the National High Technology Research and Development Program of China under Grant 2013AA01A602, in part by the Program for New Century Excellent Talents in University under Grant NCET-12-0917, in part by the Key Research Program of the Chinese Academy of Sciences under Grant KGZD-EW-T03, in part by the Australian Research Council Projects under Grant FT-130101457, Grant DP-140102164, and Grant DP-120103730, and in part by the China Post-Doctoral Science Foundation under Grant 2013M540734 and Grant 2014T70905.

C. Deng, J. Xu, and X. Gao are with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: chdeng.xd@gmail.com; xujie@stu.xidian.edu.cn; xbago.xidian@gmail.com).

K. Zhang is with the School of Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China (e-mail: kbzhang@hbeu.edu.cn).

D. Tao is with the Centre for Quantum Computation & Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, 81 Broadway Street, Ultimo, NSW 2007, Australia (e-mail: dacheng.tao@uts.edu.au).

X. Li is with the Center for Optical Imagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, China (e-mail: xuelong_li@ieee.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2015.2468069

Index Terms—Nonlocal (NL) self-similarity, structured output, super-resolution (SR), support vector regression (SVR).

NOMENCLATURE

\mathbb{R}	Real set.
\mathbb{R}_+	Positive real set.
\mathbb{N}	Positive integer set.
\otimes	Kronecker product.
$\mathbf{x} \in \mathbb{R}^d$	Vector with x_i as its i th element.
\mathbf{x}^\top	Transpose of \mathbf{x} .
$\mathbf{1}_d = [1, 1, \dots, 1]^\top \in \mathbb{R}^d$	Unit vector.
$\mathbf{0}_d = [0, 0, \dots, 0]^\top \in \mathbb{R}^d$	Zero vector.
$\mathbf{x}^\top \mathbf{z} = \sum_{k=1}^d x_k z_k$	Inner product between two vectors.
A	Linear operation.
$\mathbf{A} \in \mathbb{R}^{m \times n}$	Matrix with a_{ij} as its (i, j) th element.
\mathbf{A}^\top	Transpose of \mathbf{A} .
\mathbf{I}_m	Identity matrix of dimension $m \times m$.
$\boldsymbol{\alpha}^i \in \mathbb{R}^m$	The i th row of $\mathbf{A} \in \mathbb{R}^{m \times n}$.
$\boldsymbol{\alpha}_j \in \mathbb{R}^n$	The j th column of $\mathbf{A} \in \mathbb{R}^{m \times n}$.
$\text{Tr}(\mathbf{A}) = \sum_{i=1}^m A_{ii}$	Trace operation of \mathbf{A} .
$\text{blockdiag}(\mathbf{A}_1, \dots, \mathbf{A}_n)$	A block diagonal matrix, having $\mathbf{A}_1, \dots, \mathbf{A}_n$ as main diagonal blocks, with all other blocks being zero matrices.

I. INTRODUCTION

Due to the limitations of physical imaging systems and imaging environments, such as optical blurring, motion blurring, undersampling, and noise, it is not easy to obtain a desirable high-resolution (HR) image or an image sequence directly from imaging devices. As an effective technique to produce high-quality images using a low-cost imaging system, image super-resolution (SR) reconstruction has attracted considerable attentions and shows a great potential for many practical applications, including computer vision, video surveillance, remote-sensing images, and so on [1], [2].

More specifically, the main purpose of image SR reconstruction is to mimic the image degradation process in an imaging system and synthesize a new HR image with

more pleasing details from one or more low-resolution (LR) images. Roughly speaking, the existing SR methods can be divided into two categories according to the number of input images, namely, multiframe SR methods [3]–[8] and single-frame SR methods [9]–[16].

The classical multiframe SR methods synthesize an HR image by exploiting complementary information from multiple LR observations with subpixel displacement from the same scene [17]. There are two different types of multiframe SR approaches: 1) the frequency-domain methods [3], [8] and 2) the spatial-domain methods. The frequency-domain methods are simple and easy to implement. Nevertheless, the weaknesses of this family of methods lie in their dependence on the presumed imaging model and the preclusion from applying any prior knowledge to the reconstruction process. By contrast, the spatial-domain methods have become popular for years. The representative works include iterative back projection (IBP) [4], probability methods [e.g., ML and maximum *a posteriori* (MAP)] [5], [6], and the projection onto-convex-sets approaches [7]. The major advantages of spatial domain methods are flexible in motion and degradation models, and are beneficial to incorporate a certain prior knowledge into the reconstruction process to obtain more stable SR estimation. Nonetheless, the noticeable weaknesses of these approaches are relatively complex and computationally intensive. In particular, the SR performance may significantly degrade if there is no adequate number of LR images available for fusion.

The other group of SR approaches is single-frame based. The existing single-frame SR methods can be divided into three subcategories, i.e., interpolation-based methods, reconstruction-based methods, and example learning-based methods. Interpolation-based methods, such as bilinear, bicubic, and B-spline kernels [9], are the simplest way to increase the spatial resolution of an LR image. Although these interpolation-based methods can produce an HR image in a very effective way, the resulting images show remarkable blurring artifacts, and therefore are unsatisfactory to the practical applications. The reconstruction-based approaches typically solve the ill-posed SR problem by imposing a certain prior knowledge upon the estimated HR image. The popular priors include edge prior [10], gradient profile prior [11], total variation [12], and Bayesian prior models [13], [14]. Though the reconstruction-based methods are particularly effective for the preservation of sharp edges, they are clumsy at generating plausible details lost in the LR input.

Example learning-based single-image SR technique utilizes a training database containing millions of LR and HR image patch pairs as prior to learn the mapping relations between LR and HR images for prediction of missing high-frequency details in a given LR input [15], [16], [18]–[25]. The seminal work of the example learning-based SR methods was proposed in [15], where Markov network is applied to establish the relation model between the LR and HR patches, and the desired HR patches are obtained by the nearest neighbor (NN)-based similarity matching. Nevertheless, Freeman's method typically requires an enormous training database containing millions of LR–HR image pairs for

the generalization ability of SR performance, resulting in intensive computation cost. Chang *et al.* [16] used locally linear embedding [18] from manifold learning and neighbor embedding (NE)-based algorithm to establish the mapping relations between the LR–HR pairs. Compared with Freeman's work, the NE-based methods show better generalization ability of representing more patterns using a smaller training database. However, a fixed number of k neighbors used for embedding may lead to displeasing artifacts in the results. To alleviate the problem in the NE-based approaches, the sparse representation-based SR algorithms are proposed. Yang *et al.* [19] suggested adaptively choosing the most related neighbors based upon a jointly learnt LR–HR dictionary pair, where the sparsity penalty is employed to find linear combinations of a few members of atoms for reconstruction. Under a similar framework to Yang's work, Zeyde *et al.* [20] improved SR efficiency using Principal Component Analysis-based dimensionality reduction and orthogonal matching pursuit [21] for sparse representation. Datsenko and Elad [22] proposed a MAP framework where the prior penalized the deviation of the solution from a weighted average of k -NNs. Dong *et al.* [23] proposed a novel sparse representation-based image deblurring and the single-image SR method using adaptive sparse domain selection (ASDS) and adaptive regularization. Yu *et al.* [24] proposed to integrate learning-and reconstruction-based methods for single-image SR.

Another kind of example learning-based approaches is regression-based ones that directly establish the regression model between the LR and HR images. For example, Kim and Kwon [26] utilized support vector regression (SVR) to estimate the high-frequency details. Ni and Nguyen [27] conducted SVR in frequency domain and cast the SR as a kernel learning problem. Li and Orchard [28] adopted the interpolated image patches as input and the center pixel of residual image (i.e., the difference between the original image and the interpolated version) patches as output. The work in [29] estimated the high-frequency details based on kernel ridge regression, where the edge smoothing prior in natural image is used as a regularization term to suppress the ringing artifacts along edge regions. He and Siu [30] estimated the pixels in the HR image via Gaussian process regression (GPR) by selecting a proper covariance function. Timofte *et al.* [31] proposed an anchored neighborhood regression (ANR) approach and extended it to the global situation (GR), which used ridge regression to learn exemplar neighborhoods offline and used these neighborhoods to precompute projections that mapped LR patches to the HR counterparts. More recently, they further proposed an improved variant of ANR [32], namely, A+, which combined the best qualities of ANR and simple functions. Dong *et al.* [33] used a deep convolutional neural network (CNN) to learn an end-to-end mapping between the LR/HR images.

Although the regression-based SR models have gained a significant improvement over other existing example-based learning methods, there are still several challenging issues to be addressed. First, as a regularization framework, both SVR-based [27] and kernel-based regressors [34] are prone to excessively smooth sharp edges and produce ringing artifacts.

Second, most of the current methods typically build one or a set of regression functions between a fixed-size patch in an LR image and its corresponding pixel in the HR version, which often ignores the inherent structure information between these pairwise inputs, leading to larger reconstruction error on average. Third, most of the example-based learning methods focused on developing sophisticated image priors to recover the missing details rather than to preserve sharp edges, which will make the model complicated, and either time or space intensive.

As stated in [35], an accurate reconstruction constraint combined with a simple image prior can produce good SR results. In this paper, we propose a novel single-image SR framework by exploiting structure output regression machine (SORM) to establish the mapping relation between LR and HR image patches. In contrast to those traditional regression-based SR methods that predict the HR image pixel-by-pixel, we utilize the SORM to learn the relations between LR and HR images in patchwise style, which can not only preserve the fine-grained structure information well but also effectively alleviate the ringing artifacts along edges. Meanwhile, considering that there exist many similar patterns repeated in natural images, we encapsulate nonlocal (NL) self-similarity constraint on the estimated HR image as image prior into a unified reconstruction framework, which is beneficial to preserve sharpness of edges and suppress noise. Extensive experiments demonstrate that our approach gains superior SR results in terms of both quantitative measure and visual perception quality. Moreover, our method shows lower time complexity in both the training and testing procedures.

The remainder of this paper is organized as follows. Section II introduces the related work that is important to the proposed method. Section III introduces problem definition. Then, we describe and analyze our algorithm and model in Section IV. The experimental results are given in Section V. Finally, the conclusion is drawn in Section VI.

II. RELATED WORK

The traditional multioutput SVR (MSVR) models each output individually. By assembling multiple single-output regression models, the multioutput regression model can be accomplished. However, this method ignores correlations between different output components, and thus reduces the regression accuracy.

A good MSVR should consider not only the underlying correlations between the inputs and the corresponding outputs, but also the correlations between the outputs [36], [37]. In order to take advantage of the possible correlations to improve the regression accuracy, constructing multioutput regression model directly by cumulating the objective functions is a natural thought. Brudnak [38] proposed a vector-valued extension of SVR case by extending the loss function, estimator, and regularization from the scalar case to a vector one. Vazquez and Walter [39] extended SVR to multioutput systems that made it possible to take advantage of the possible correlations between the outputs to improve the quality of the predictions. Pérez-Cruz *et al.* [40] proposed an MSVR that used a cost function with a hyperspherical

insensitive zone, capable of obtaining better predictions than using an support vector machine (SVM) independently for each dimension. Sánchez-fernández *et al.* [41] used MSVR to discover the dependence between transmitted and received signals in a multiple-input multiple-output system. To further keep the corresponding relation between input and output, Vedaldi *et al.* [42] presented a generalization of structured output SVM regressor by incorporating equivariance and invariance into a convex training procedure, achieving the same performance while using fewer samples.

Inspired by multitask learning methods, An *et al.* [43] and Xu *et al.* [44], [45] extended the least-square SVR (LS-SVR) to the multiple-output LS-SVR (MLS-SVR), which took full consideration into the circumstance that each submodel may affected the other submodels during the training process. In particular, compared with the standard SVMs, the MLS-SVR adopts the quadratic component as the optimizing index, which reduces one tuning parameter. As a result, it can improve computational efficiency as well as reconstruction quality. This paper introduces MLS-SVR to single-image SR and encapsulates an NL regularization term into structure-output regression model.

III. PROBLEM DEFINITION

Let us first recall a common degradation model of single-image SR problem. Assuming that the observed LR image \mathbf{X} results from warping, blurring, and downsampling operators performed on an unknown HR image \mathbf{Y} , we can represent the observation model as

$$\mathbf{X} = D\mathbf{H}\mathbf{Y} + \eta \quad (1)$$

where H is the blurring operator, D is the downsampling operator, and η is the additive noise. The SR reconstruction problem is to estimate the target HR version \mathbf{Y} from the observed image \mathbf{X} . In the regression-based SR methods, an LR image is spliced into an arranged set of image patches, and then, SR is conducted on each patch. Given a set of LR and HR patch pairs $\{(\mathbf{p}_i^l, \mathbf{p}_i^h), i = 1, \dots, N\}$ extracted from the example images, the corresponding HR patches of the resultant HR image are estimated based on the co-occurrence relation between patch pairs $(\mathbf{p}_i^l, \mathbf{p}_i^h)$ in the training procedure.

As a common scenario in the regression-based SR, one LR pixel corresponds to four HR pixels in the case of $2 \times$ SR magnification. As an example, Ni and Nguyen [27] took the LR image patches removed the center pixel as input and the corresponding HR image patches as output, and then divided the SR problem into several single-output regression problems, which only built the relation between an input LR patch and a pixel of an output HR patch, and thus isolated the pixel in the HR patch with its spatial neighbors. Therefore, a fatal defect in this method is their serious ringing artifacts on the edge of the reconstructed image. To address this problem, a natural idea is to preserve the spatial relation among the reconstructed pixels in an HR patch as much as possible. In this paper, we utilize a novel regression model, i.e., SORM, to build the accurate spatial relation between LR patches and their corresponding HR patches. Therefore,

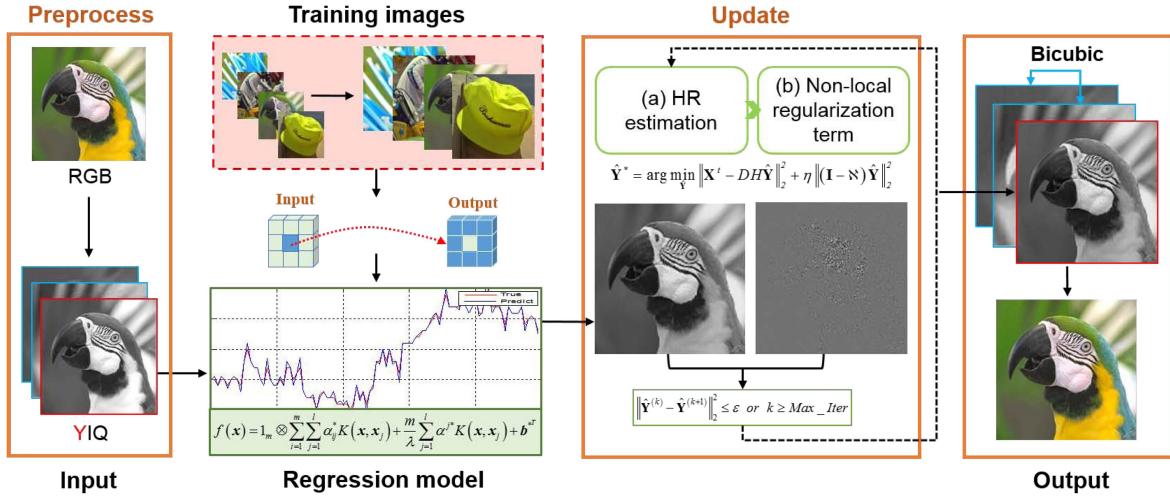


Fig. 1. Framework of the proposed SR method.

the key task of the proposed method is to solve a multivariate and multidimensional problem for this regression model.

IV. STRUCTURED OUTPUT REGRESSION MACHINE

In this section, we first describe a popular single-output regression model, i.e., LS-SVR, and then extend it as in [43]–[45] to the situation of structured multiple outputs, namely, SORM. Then, the NL self-similarity prior is introduced into the proposed SORM-based SR framework as a regularization term to further improve the reconstruction quality. The diagram of the proposed single-image SR method is shown in Fig. 1.

A. Least-Squares Support Vector Regression

Given a training data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^l$, where $\mathbf{x}_i \in \mathbb{R}^d$ is an input vector with dimensionality d , and $y_i \in \mathbb{R}$ is the corresponding output with dimensionality one. A nonlinear mapping $\varphi(\cdot)$ is used to map the data set from input space to high-dimensional feature space, such that the nonlinear fitting problem is turned into a linear problem [46]. The linear regression function in high-dimensional feature space is formulated as

$$f(\mathbf{x}) = \mathbf{w}^\top \varphi(\mathbf{x}) + b \quad (2)$$

where $\mathbf{w} \in \mathbb{R}^{n_h}$, $\varphi(\mathbf{x}) \in \mathbb{R}^{n_h}$, and b is a threshold.

According to the structural risk minimization principle and also considering the function complexity and fitting error, the LS-SVR problem can be expressed as a constrained optimization problem as follows:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} J(\mathbf{w}, \xi) &= \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^l \xi_i^2 \\ \text{s.t. } y_i &= \mathbf{w}^\top \varphi(\mathbf{x}_i) + b + \xi_i, \quad i = 1, 2, \dots, l \end{aligned} \quad (3)$$

where ξ_i represents the fitting error and γ is a regularization parameter. By introducing Lagrange multipliers,

the Lagrangian of (3) is defined as

$$L(\mathbf{w}, b, \xi, \alpha) = J(\mathbf{w}, \xi) - \sum_{i=1}^l \alpha_i [\mathbf{w}^\top \varphi(\mathbf{x}_i) + b + \xi_i - y_i] \quad (4)$$

where $\alpha_i \geq 0 (i = 1, \dots, l)$.

According to Karush–Kuhn–Tucker (KKT) conditions, the partial derivatives of (4) are taken as

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^l \alpha_i \varphi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^l \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma \xi_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \mathbf{w}^\top \varphi(\mathbf{x}_i) + b + \xi_i - y_i = 0. \end{cases} \quad (5)$$

On the basis of Mercer condition, we can define the kernel function as $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^\top \varphi(\mathbf{x}_j)$. Eliminating ξ_i and \mathbf{w} in (5), we obtain the following set of linear equations:

$$\begin{bmatrix} 0 & \mathbf{1}_l^\top \\ \mathbf{1}_l & \mathbf{K} + \frac{1}{\gamma} \mathbf{I}_l \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}. \quad (6)$$

Solving (6), we can obtain $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^\top$ and b^* . Thus, the predictive model of LS-SVR can be depicted as

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}^{*\top} \varphi(\mathbf{x}) + b^* = \sum_{i=1}^l \alpha_i^* \varphi(\mathbf{x}_i)^\top \varphi(\mathbf{x}) + b^* \\ &= \sum_{i=1}^l \alpha_i^* K(\mathbf{x}_i, \mathbf{x}) + b^*. \end{aligned} \quad (7)$$

It follows that one can solve the single-output LS-SVR only using inner product between inputs. Compared with the standard SVM, the LS-SVR only needs to determine the shape parameter and punish the coefficient of the kernel function without the values of insensitive loss function, which simplifies the calculation and facilitates practical applications.

B. Structured Output Regression Machine

It is easy to extend the single-output regression model (i.e., LS-SVR) to multioutput situation, also called SORM in here. Given a set of independent and identically distributed samples, i.e., $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^l$, $\mathbf{x}_i \in \mathbb{R}^d$, $\mathbf{y}_i \in \mathbb{R}^m$, the SORM model aims to predict an output vector \mathbf{y}_i with an input vector \mathbf{x}_i [44], [45].

Mathematically, the SORM model can be cast as the following optimization problem:

$$\begin{aligned} \min J(\mathbf{w}_i, \boldsymbol{\xi}_i) &= \frac{1}{2m} \sum_{i=1}^m \mathbf{w}_i^\top \mathbf{w}_i + \frac{\gamma}{2} \sum_{i=1}^m \boldsymbol{\xi}_i^2 \\ \text{s.t. } \mathbf{y}_i &= \mathbf{Z}^\top \mathbf{w}_i + \mathbf{b}_i \mathbf{1}_l + \boldsymbol{\xi}_i, \quad i \in \mathbb{N}_m \end{aligned} \quad (8)$$

where \mathbf{y}_i , \mathbf{w}_i , and $\boldsymbol{\xi}_i \in \mathbb{R}^l$, $\mathbf{Z} = (\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_l)) \in \mathbb{R}^{n_h \times l}$.

For convenience, (8) is rewritten in a matrix form as

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{b}} J(\mathbf{W}, \Xi) &= \frac{1}{2} \text{Tr}(\mathbf{W}^\top \mathbf{W}) + \frac{\gamma}{2} \text{Tr}(\Xi^\top \Xi) \\ \text{s.t. } \mathbf{Y} &= \mathbf{Z}^\top \mathbf{W} + \mathbf{1}_l \otimes \mathbf{b}^\top + \Xi \end{aligned} \quad (9)$$

where

$$\begin{aligned} \mathbf{Z} &= (\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), \dots, \varphi(\mathbf{x}_l)) \in \mathbb{R}^{n_h \times l}; \\ \mathbf{W} &= (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m) \in \mathbb{R}^{n_h \times m}; \\ \Xi &= (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m) \in \mathbb{R}_+^{l \times m}; \\ \mathbf{b} &= (b_1, b_2, \dots, b_m)^\top \in \mathbb{R}^m. \end{aligned}$$

It is not difficult to see that this problem is equivalent to m optimization problems, which is similar to the problem in (3) [44]. But it is much more efficient to solve (9) directly than to solve (3) m times, since they all share the same matrix $\mathbf{H} = \mathbf{K} + \gamma^{-1} \mathbf{I}_l \in \mathbb{R}^{l \times l}$, the inverse matrix of which just needs to be calculated only once with some off-the-shelf methods, such as Cholesky decomposition, conjugate gradient, or Singular Value Decomposition.

We assume that all $\mathbf{w}_i \in \mathbb{R}^{n_h} (\forall i \in \mathbb{N}_m)$ can be written as $\mathbf{w}_i = \mathbf{w}_0 + \mathbf{v}_i$, where $\mathbf{v}_i \in \mathbb{R}^{n_h} (i \in \mathbb{N}_m)$ are small when the different outputs are similar to each other, otherwise the mean vector $\mathbf{w}_0 \in \mathbb{R}^{n_h}$ are small. To solve $\mathbf{w}_0 \in \mathbb{R}^{n_h}$, $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) \in \mathbb{R}^{n_h \times m}$ and $\mathbf{b} = (b_1, b_2, \dots, b_m)^\top \in \mathbb{R}^m$ simultaneously, we rewrite the constrained minimization problem in (9) as follows:

$$\begin{aligned} \min_{\mathbf{W}_0, \mathbf{V}, \mathbf{b}} J(\mathbf{W}_0, \mathbf{V}, \Xi) &= \frac{1}{2} \mathbf{W}_0^\top \mathbf{W}_0 + \frac{\lambda}{2m} \text{Tr}(\mathbf{V}^\top \mathbf{V}) + \frac{\gamma}{2} \text{Tr}(\Xi^\top \Xi) \\ \text{s.t. } \mathbf{Y} &= \mathbf{Z}^\top \mathbf{W} + \mathbf{1}_l \otimes \mathbf{b}^\top + \Xi \end{aligned} \quad (10)$$

where

$$\begin{aligned} \Xi &= (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m) \in \mathbb{R}^{l \times m}; \\ \mathbf{W} &= (\mathbf{w}_0 + \mathbf{v}_1, \mathbf{w}_0 + \mathbf{v}_2, \dots, \mathbf{w}_0 + \mathbf{v}_m) \in \mathbb{R}^{n_h \times m}; \\ \mathbf{Z} &= (\varphi(\mathbf{x}_1), \varphi(\mathbf{x}_2), \dots, \varphi(\mathbf{x}_l)) \in \mathbb{R}^{n_h \times l}; \\ \lambda \text{ and } \gamma &\text{ are two regularization parameters.} \end{aligned}$$

The Lagrangian function of (11) is reformulated as

$$\begin{aligned} L(\mathbf{W}_0, \mathbf{V}, \mathbf{b}, \Xi, \mathbf{A}) &= J(\mathbf{W}_0, \mathbf{V}, \Xi) - \text{Tr}(\mathbf{A}^\top (\mathbf{Z}^\top \mathbf{W} + \mathbf{1}_l \otimes \mathbf{b}^\top + \Xi - \mathbf{Y})) \end{aligned} \quad (11)$$

where $\mathbf{A} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_m)^\top \in \mathbb{R}^{l \times m}$ is a matrix consisting of Lagrange multipliers. According to the KKT conditions, the partial derivatives of (11) are taken as

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial \mathbf{w}_0} = 0 \rightarrow \mathbf{w}_0 = \sum_{i=1}^m \mathbf{Z} \boldsymbol{\alpha}_i \\ \frac{\partial L}{\partial \mathbf{V}} = 0 \rightarrow \mathbf{V} = \frac{m}{\lambda} \mathbf{Z} \mathbf{A} \\ \frac{\partial L}{\partial \mathbf{b}} = 0 \rightarrow \mathbf{A}^\top \mathbf{1}_l = \mathbf{0}_l \\ \frac{\partial L}{\partial \Xi} = 0 \rightarrow \mathbf{A} = \gamma \Xi \\ \frac{\partial L}{\partial \mathbf{A}} = 0 \rightarrow \mathbf{Z}^\top \mathbf{W} + \mathbf{1}_l \otimes \mathbf{b}^\top + \Xi - \mathbf{Y} = \mathbf{0}_{l \times m}. \end{array} \right. \quad (12)$$

By eliminating \mathbf{W} and Ξ , one can obtain the following linear system:

$$\begin{bmatrix} \mathbf{0}_{ml \times m} & \mathbb{C}^\top \\ \mathbb{C} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_m \\ \mathbf{y} \end{bmatrix} \quad (13)$$

where

$$\begin{aligned} \mathbb{C} &= \text{blockdiag}(\overbrace{\mathbf{1}_l, \dots, \mathbf{1}_l}^m) \in \mathbb{R}^{ml \times m}; \\ \mathbf{H} &= \Omega + \gamma^{-1} \mathbf{I}_{ml} + (m/\lambda) \mathbf{Q} \in \mathbb{R}^{ml \times ml}; \\ \Omega &= \mathbf{I}_m \otimes \mathbf{K}^\top \in \mathbb{R}^{ml \times ml}; \\ \mathbf{Q} &= \text{blockdiag}(\overbrace{\mathbf{K}, \dots, \mathbf{K}}^m) \in \mathbb{R}^{ml \times ml}; \\ \mathbf{K} &= \mathbf{Z}^\top \mathbf{Z} \in \mathbb{R}^{l \times l}; \\ \boldsymbol{\alpha} &= (\boldsymbol{\alpha}_1^\top, \dots, \boldsymbol{\alpha}_m^\top)^\top \in \mathbb{R}^{ml}; \\ \mathbf{y} &= (\mathbf{y}_1^\top, \dots, \mathbf{y}_m^\top)^\top \in \mathbb{R}^{ml}. \end{aligned}$$

Thus, the linear system in (13) consists of $(l+1) \times m$ equations.

Solving (13) obtains $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^{*\top}, \dots, \boldsymbol{\alpha}_m^{*\top})^\top$ and \mathbf{b}^* . Thus, the predictive model of the SORM can be expressed as

$$\begin{aligned} f(\mathbf{x}) &= \varphi(\mathbf{x})^\top \mathbf{W}^* + \mathbf{b}^{*\top} \\ &= \varphi(\mathbf{x})^\top (\mathbf{1}_m \otimes \mathbf{w}_0^*) + \varphi(\mathbf{x})^\top \mathbf{V}^* + \mathbf{b}^{*\top} \\ &= \varphi(\mathbf{x})^\top \left(\mathbf{1}_m \otimes \sum_{i=1}^m \mathbf{Z} \boldsymbol{\alpha}_i^* \right) + \frac{m}{\lambda} \varphi(\mathbf{x})^\top \mathbf{Z} \mathbf{A}^* + \mathbf{b}^{*\top} \\ &= \mathbf{1}_m \otimes \sum_{i=1}^m \sum_{j=1}^l a_{ij}^* K(\mathbf{x}, \mathbf{x}_j) + \frac{m}{\lambda} \sum_{j=1}^l \boldsymbol{\alpha}^{j*} K(\mathbf{x}, \mathbf{x}_j) + \mathbf{b}^{*\top}. \end{aligned} \quad (14)$$

Unlike single-output LS-SVR model, the SORM aims at solving the multivariate and multidimensional problems, i.e., the input and output values of the system are all multidimensional. It takes full consideration into the circumstance that each subproblem may impact on other subproblems during the training process. That is to say, the SORM model considers the underlying (potentially nonlinear) cross relatedness among different outputs, and therefore is superior to LS-SVR. Moreover, for our proposed regression-based SR problem, we need this property of the SORM model just right.

C. Nonlocal Self-Similarity Constraint

Self-similarity, one of the most useful properties of natural images, means that the small image patches tend to repeat themselves many times, which can be exploited as useful prior knowledge for various image restoration tasks.

Self-similarity in a natural image within the same scale can be well described by NL means (NLMs) [47]. The NLMs assume that the central pixel of a local patch can be approximated by the convex combination of the central pixels of their similar patches (i.e., similar pixels) within a large neighborhood range. Mathematically, the NLMs can be formulated as

$$y_i = \frac{\sum_{j \in P(y_i)} \omega_{ij} y_j}{\sum_{j \in P(y_i)} \omega_{ij}} \quad (15)$$

where $P(y_i)$ represents the index set, which is formed by the pixel coordinates of similar pixels, and ω_{ij} is the similarity weight between the local patch y_i at the location y_i and the local patch y_j at the location y_j , which is calculated by

$$\omega_{ij} = \left(1 + \frac{\|R_{y_i}\mathbf{Y} - R_{y_i}\mathbf{Y}\|_G^2}{h_n^2} \right)^{-1} \quad (16)$$

where h_n is a global smoothing parameter controlling the decay rate of exponential expression, G is a kernel matrix that assigns the larger weights to the pixels close to the center pixel within the target image patch, and R_{y_i} is an operator to extract the i th local patch y_i in \mathbf{Y} . Dong *et al.* [48] proposed long ago to adaptively incorporate the NL information into the IBP process so that the reconstruction errors could be reduced. The weight computation is the most time consuming part when generating the restored image. According to Darbon's work [49], we adopt this acceleration algorithm to efficiently calculate the weights ω_{ij} , which will yields $O(2^d)$ computational cost for d -dimensional image. Formally, they assume images are defined over a discrete regular grid \mathbb{D} of dimension d and cardinality $|\mathbb{D}|$.

The NL regularization term can be written as

$$\hat{\mathbf{Y}} = \operatorname{argmin}_{\mathbf{Y}} \left\{ \sum_{i \in X} \|y_i - c_i^N \cdot S_i\|_2^2 \right\} \quad (17)$$

where S_i is a column vector, which is arranged by similar pixels in the order, c_i^N is composed of the corresponding NL weights, and y_i is the i th pixel of the HR image to be estimated. The formula can be rewritten as

$$\hat{\mathbf{Y}} = \operatorname{argmin}_{\mathbf{Y}} \|(\mathbf{I} - \mathbf{N})\mathbf{Y}\|_2^2 \quad (18)$$

where \mathbf{I} is the identity matrix and

$$\mathbf{N}(i, j) = \begin{cases} c_{ij}^N, & j \in S_i \\ 0, & j \notin S_i. \end{cases}$$

D. Summary of the Algorithm

Sections IV-A–IV-C describe the structured output least-squares regression and NL self-similarity. We first discuss how

to combine the SORM model and the self-similarity constraint, and then apply them to the image SR in this section.

We first elaborate the input and the output of the proposed regression-based SR problem. As discussed in Section III, a set of LR and HR patch pairs $\{(\mathbf{p}_i^l, \mathbf{p}_i^h), i = 1, \dots, N\}$ is extracted from the example images. We assume that the number of training and testing patches is l_s and l_t , respectively.

Then, we use bicubic to interpolate the LR image to the size of desired HR image. Both LR and HR images are divided into 3×3 patches, upon which the relations between LR and HR patches are built. In the reconstruction phase, the patches are extracted with one pixel overlapping. To capture not only the underlying relations between the structured inputs and the corresponding outputs but also the relations between the outputs, we view this SR task as multiple-output regression problems and adopt the proposed SORM model to solve it. The algorithm process is divided into two phases, and the specific procedures are as follows.

1) *Training Phase*: We use the training data sets $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{l_s}$ and solve (11) to obtain the parameters of the SORM model, $\boldsymbol{\alpha}^* = (\boldsymbol{\alpha}_1^{*\top}, \dots, \boldsymbol{\alpha}_l^{*\top})^\top$ and \mathbf{b}^* . Thus, the predict model can be written as

$$f(\mathbf{x}) = \varphi(\mathbf{x})^\top \mathbf{W}^* + \mathbf{b}^{*\top}. \quad (19)$$

2) *Reconstruction Phase*: Because of the noise and other reasons, the results of the predict model may not satisfy the reconstruction constraint in (1) exactly. We eliminate this discrepancy by projecting $\hat{\mathbf{Y}} = \{\hat{y}_1, \dots, \hat{y}_{l_t}\}$ onto the solution space of $D\mathbf{H}\mathbf{Y} = \mathbf{X}$, computing

$$\hat{\mathbf{Y}}^* = \operatorname{argmin}_{\mathbf{Y}} \|\mathbf{Y} - \hat{\mathbf{Y}}\|, \quad \text{s.t. } D\mathbf{H}\mathbf{Y} = \mathbf{X}. \quad (20)$$

By incorporating the NL similarity regularization term into the SORM model in (14), the SR problem can be written as

$$\begin{aligned} \hat{\mathbf{Y}}^* &= \operatorname{argmin}_{\hat{\mathbf{Y}}} \|\mathbf{X}^t - D\mathbf{H}\hat{\mathbf{Y}}\|_2^2 + \eta \|(\mathbf{I} - \mathbf{N})\hat{\mathbf{Y}}\|_2^2 \\ \text{s.t. } \hat{\mathbf{Y}} &= (\mathbf{Z}^t)^\top \mathbf{W}^* + \mathbf{1}_l \otimes \mathbf{b}^{*\top} \end{aligned} \quad (21)$$

where $\mathbf{X}^t = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{l_t}]^\top$ is the LR test data sets.

To facilitate implementation, we seek a local optimal solution through a two-stage scheme.

1) In the first stage, we input the test data sets $\{\mathbf{x}_i^t, \mathbf{y}_i^t\}_{i=1}^{l_t}$ to the predict model directly

$$\begin{aligned} \hat{\mathbf{y}} &= \varphi(\mathbf{x}^t)^\top \mathbf{W}^* + \mathbf{b}^{*\top} = \mathbf{1}_l \otimes \sum_{i=1}^m \sum_{j=1}^l \alpha_{ij}^* K(\mathbf{x}^t, \mathbf{x}_j^t) \\ &\quad + \frac{m}{\lambda} \sum_{j=1}^l \boldsymbol{\alpha}^{j*} K(\mathbf{x}^t, \mathbf{x}_j^t) + \mathbf{b}^{*\top} \end{aligned} \quad (22)$$

where $\{\mathbf{y}_i^t\}_{i=1}^{l_t}$ are the actual outputs of test data sets $\{\mathbf{x}_i^t, \mathbf{y}_i^t\}_{i=1}^{l_t}$ and only used in the final objective quality assessing. After finishing this stage, we obtain an initial estimation of the HR image $\hat{\mathbf{Y}} = [\hat{y}_1, \dots, \hat{y}_{l_t}]^\top$.

Algorithm 1 SORM Based Single Image SR

- 1: Input training set $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^l$, we compute $\boldsymbol{\alpha}^*$ and \mathbf{b}^* using Eq. (13);
- 2: Estimate $\hat{\mathbf{y}}$ by using test set $\{\mathbf{x}_i^t, \mathbf{y}_i^t\}_{i=1}^l$ and Eq. (14);
- 3: Calculate the non-local weight $\mathbf{\omega}_i$ for each local patch $\hat{\mathbf{y}}_i$ by using Eq. (16);
- 4: Initialize $\mathbf{\aleph}$ with the selected non-local weights;
- 5: Preset η , Θ , ϵ , and the maximal iteration number, Max_Iter ;
- 6: Set $k = 0$;
- 7: Iterate on k until $\|\hat{\mathbf{Y}}^{(k)} - \hat{\mathbf{Y}}^{(k+1)}\|_2^2 \leq \epsilon$ or $k \geq Max_Iter$ is satisfied;
- 8: $\hat{\mathbf{Y}}^{(k+\frac{1}{2})} = \hat{\mathbf{Y}}^{(k)} + \tau \{H^\top D^\top (\mathbf{X}^t - DH\hat{\mathbf{Y}}^{(k)}) - \eta(\mathbf{I} - \mathbf{\aleph})^\top (\mathbf{I} - \mathbf{\aleph})\hat{\mathbf{Y}}^{(k)}\}$,
where D and H are degradation operators;
- 9: Compute $\hat{\mathbf{Y}}^{(k+1)}$ using Eq. (14);
- 10: If $mod(k, \Theta) = 0$, update the matrix $\mathbf{\aleph}$ by using the improved estimation $\hat{\mathbf{Y}}^{(k+1)}$;



Fig. 2. Some examples of test images.

- 2) In the second stage, we solve the following MAP problem as:

$$\hat{\mathbf{Y}}^* = \underset{\hat{\mathbf{Y}}}{\operatorname{argmin}} \|\mathbf{X}^t - DH\hat{\mathbf{Y}}\|_2^2 + \eta \|(\mathbf{I} - \mathbf{\aleph})\hat{\mathbf{Y}}\|_2^2. \quad (23)$$

This problem can be solved by the gradient descent rule

$$\hat{\mathbf{Y}}^{(k+\frac{1}{2})} = \hat{\mathbf{Y}}^{(k)} + \tau \{H^\top D^\top (\mathbf{X}^t - DH\hat{\mathbf{Y}}^{(k)}) - \eta(\mathbf{I} - \mathbf{\aleph})^\top (\mathbf{I} - \mathbf{\aleph})\hat{\mathbf{Y}}^{(k)}\} \quad (24)$$

where k is the iteration number, τ is the step size for gradient descent, and $\hat{\mathbf{Y}}$ is an initial estimation to the final HR output $\hat{\mathbf{Y}}^*$.

We use the initial HR image $\hat{\mathbf{Y}}$ predicted by the SORM model as the initial solution of the aforementioned MAP problem. In each step, we perform the above two stages alternately to obtain an HR image estimation. The specific procedures are summarized in Algorithm 1.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In order to verify the effectiveness of the proposed single-image SR approach, we carry out $2\times$ and $3\times$ SR reconstruction on ten test images shown in Fig. 2. We compare our method with several representative example learning-based SR methods, including the traditional GPR [30], GR [31], ANR [31], A+ [32], ASDS [23], and CNN [33]. All the compared results are reproduced by the MATLAB codes

downloaded from the authors' websites. Moreover, we retrain each method using the same blurring and downsampling operations.

A. Experimental Settings

To simulate the image degradation process, the LR images are generated by first applying a truncated 7×7 Gaussian kernel of standard deviation 1.1 to the original image and downsampling by a factor of 2 or 3. Since the human visual system is more sensitive to the luminance component of the image than to the chrominance ones, we convert the color image from the RGB color space to the YIQ color space and only perform SR reconstruction on the luminance component, and the other two components are simply magnified by bicubic. We use bicubic to interpolate the LR image to the size of HR image first. The sizes of the LR and HR image patches are both 3×3 .

For example learning-based image SR methods, training images have a certain impact on the quality of the reconstructed images. For a fair comparison, we select a large number of HR images similar to [31] and randomly select 50000 LR and HR patch pairs as training samples. As aforementioned in Section IV-D, we construct the relations between LR and HR image patches. Here, we first use A Library for Support Vector Machines to determine support vector sample set S by establishing nine single-output regression models from LR patches to one HR pixel and randomly choose 500 training samples from nonsupport vector sample set (i.e., $50000-S$). We use these 500 samples to train the proposed SORM model. In Section V-B, we will detail the major reason why we only use nonsupport vector sample set as the training samples.

B. Parameters Selection

In this paper, the linear and radial basis function (RBF) kernel functions are adopted, i.e., $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$ and $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-p\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, $p > 0$, respectively. In training phase, the grid search is used to determine appropriate parameters. Let $\gamma \in \{2^{-5}, 2^{-3}, \dots, 2^{15}\}$, $\lambda \in \{2^{-10}, 2^{-8}, \dots, 2^{10}\}$, and $p \in \{2^{-15}, 2^{-13}, \dots, 2^3\}$. For all possible combinations (γ and λ) with linear kernel function or (γ , λ , and p) with RBF kernel function, the average relative error $\delta = (1/l) \sum_{i=1}^l (\|\mathbf{y}_i - \hat{\mathbf{y}}_i\|/\|\mathbf{y}_i\|)$ is calculated using leave-one-out procedure, where \mathbf{y}_i and $\hat{\mathbf{y}}_i$ are the actual and predicted outputs, respectively. Thus, an optimal triple (γ^*, λ^*) or $(\gamma^*, \lambda^*, p^*)$ can be obtained. With the optimal parameters, we can get the parameters ($\boldsymbol{\alpha}^*$ and \mathbf{b}^*) of the SORM model on the basis of Section IV-B, and thus, the predictive model is obtained.

1) Effect of Kernel Functions: The type of kernel implicitly determines the nonlinear mapping from input space to high-dimensional feature space, which is very important to the performance of SVMs. In order to analyze the impact of the kernel function on the SORM model, we compare the quality of the recovered images under different kernels. We choose linear and RBF kernel, respectively. The averaged results of PSNR, SSIM, feature similarity (FSIM) [50], and root-mean-squared error (RMSE) of ten test images are

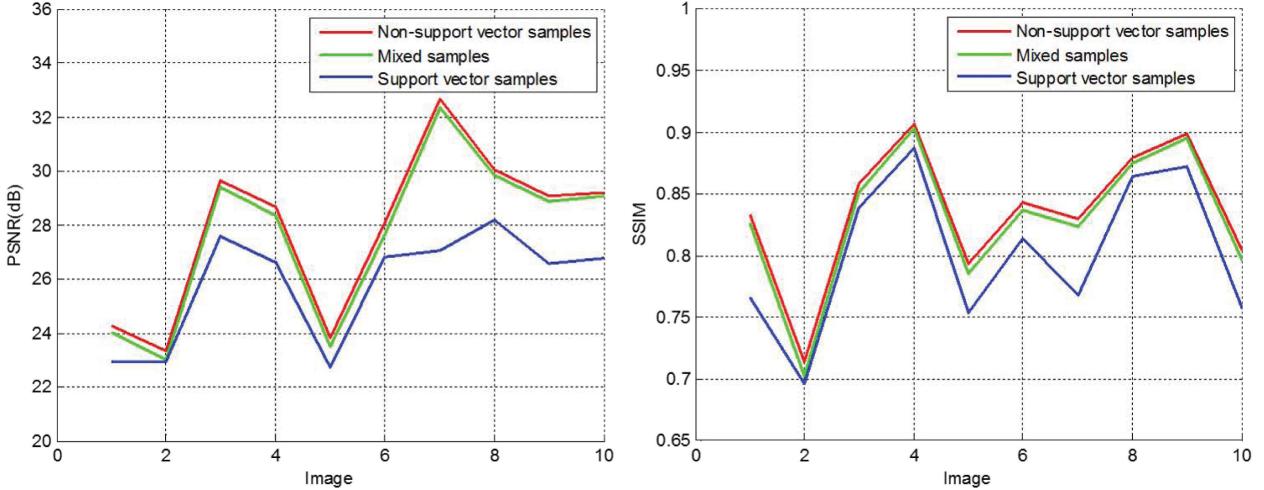


Fig. 3. PSNR and SSIM results of ten test images with various training samples under a magnification factor of 3.

TABLE I
QUANTITATIVE ASSESSING (3×) OF DIFFERENT KERNELS

Methods	Image Quality Metrics				Elapsed(s)
	PSNR	SSIM	FSIM	RMSE	
Linear	27.87	0.836	0.889	10.73	1.2
RBF	27.85	0.833	0.888	10.78	4.7

compared in Table I. In order to only measure the influence by the kernels, the quantitative assessment results in Table I are obtained without applying NL regularization term. As shown in Table I, the difference between the results under different kernels is trivial. However, using RBF kernel needs to spend a lot of time, nearly four times more than the training time of the linear kernel. Therefore, we choose to use linear kernel function in the following experiments.

2) *Effect of Training Samples*: As aforementioned, we use nonsupport vector samples as the training data to learn the SORM model. We believe that the regression model determined by support vector samples is suitable for the reconstruction of most common image patches rather than the representative patches. However, in image SR, the representative image patches are difficult to recover well, which just corresponds to the nonsupport vector samples. Therefore, it will become more efficient if we can find out these nonsupport vector samples and use them to train the SORM model.

To verify the above observation, we employ 500 support vector samples, 500 mixed samples, and 500 nonsupport vector samples to train the SORM model, respectively, and compare the PSNR and SSIM on ten test images, as shown in Fig. 3. The average PSNR and SSIM obtained from nonsupport vector samples are best, which further proves the reasonableness of our choice of nonsupport vector samples as training data.

3) *Effectiveness of NLMs*: To validate the effectiveness of NL self-similarity constraint, we compare the SR performance with varying iterations of NL regularization terms, as shown in Fig. 4. The fluctuation in Fig. 4 indicates that the PSNR and RMSE achieve the best and keep stable at nearly ten iteration times. Meanwhile, the SSIM and the FSIM are also in the same condition. Therefore, we choose 30 iteration

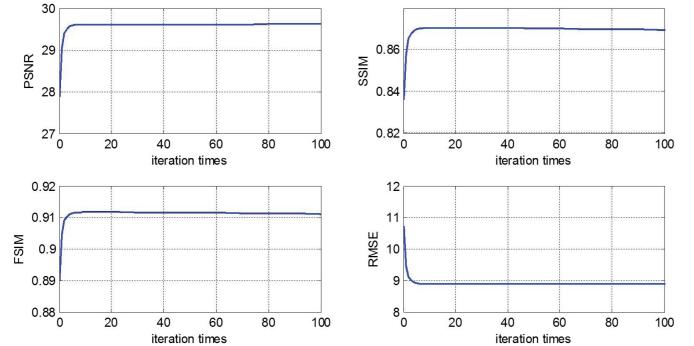


Fig. 4. Average PSNR, SSIM, FSIM, and RMSE of ten test images (3×) varying with iteration times.

times to balance objective evaluation metrics and elapsed time at 3× magnification.

C. Experimental Results

Objectively, we use PSNR and SSIM to evaluate the performance of different SR methods. The PSNR and SSIM values of 2× and 3× reconstructed HR images are shown in Tables II and III, respectively. It can be observed that the proposed method constantly outperforms all of the comparison methods in the most cases of both 2× and 3× magnification factors. All the reconstructed results of SORM + NL are better than those of SORM, which indicates the effectiveness of the NL regularization term.

To further evaluate the SR performances among these methods, we visually assess the SR results obtained from different methods, as shown in Figs. 5–8. In particular, Figs. 5 and 6 are the results under 2× magnification factor, and Figs. 7 and 8 are the results under 3× magnification factor, respectively. To compare the capabilities of recovering details among these SR methods, we use bicubic interpolation to enlarge the region of interest by a factor of 2, as shown in Figs. 5–8 (red insets).

As shown in Figs. 5(c) and 6(c), the GPR method [30] cannot recover high-frequency details effectively, and therefore, the reconstructed images are very blurry. Similarly, there are many unpleasing ringing effects in

TABLE II
PSNR (dB) AND SSIM RESULTS (LUMINANCE COMPONENTS)
OF 2 \times RECONSTRUCTED HR IMAGES

Image	GPR [30]		GR [31]		ANR [31]		ASDS [23]		CNN [33]		Proposed			
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	SORM		SORM+NL	
											PSNR	SSIM	PSNR	SSIM
Cameraman	21.32	0.733	26.46	0.870	28.20	0.893	27.46	0.898	27.70	0.898	25.81	0.871	28.44	0.908
Fence	20.44	0.520	23.97	0.777	25.70	0.788	26.42	0.813	25.99	0.806	24.94	0.787	26.85	0.832
House	25.79	0.781	30.76	0.885	35.14	0.913	34.67	0.911	34.38	0.907	31.72	0.895	34.91	0.914
Parrots	25.47	0.841	29.17	0.914	32.23	0.941	33.65	0.949	33.13	0.951	30.37	0.931	33.96	0.955
Bike	21.18	0.649	24.67	0.843	26.94	0.881	28.13	0.909	28.06	0.905	25.33	0.849	28.64	0.918
Flower	24.73	0.714	28.20	0.869	31.57	0.912	32.72	0.930	32.45	0.926	29.75	0.887	32.83	0.933
Girl	29.87	0.751	32.76	0.864	35.06	0.873	35.43	0.882	35.48	0.886	34.12	0.865	35.68	0.891
Hat	27.38	0.813	30.82	0.889	33.32	0.920	34.05	0.927	34.09	0.931	31.55	0.907	34.31	0.935
Parents	25.30	0.799	29.24	0.907	33.53	0.945	34.35	0.952	34.39	0.954	31.12	0.930	34.86	0.958
Raccoon	26.58	0.670	29.62	0.869	31.41	0.869	32.63	0.898	32.42	0.897	30.72	0.858	32.89	0.907
Average	24.81	0.727	28.57	0.869	31.31	0.894	31.95	0.907	31.81	0.906	29.54	0.878	32.34	0.915

TABLE III
PSNR (dB) AND SSIM RESULTS (LUMINANCE COMPONENTS) OF 3 \times RECONSTRUCTED HR IMAGES

Image	GPR [30]		ANR [31]		A+ [32]		ASDS [23]		CNN [33]		Proposed			
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	SORM		SORM+NL	
											PSNR	SSIM	PSNR	SSIM
Cameraman	20.96	0.712	25.60	0.826	26.20	0.843	24.43	0.818	24.48	0.817	24.27	0.833	25.77	0.864
Fence	20.42	0.491	22.26	0.620	22.35	0.638	22.22	0.632	22.18	0.618	23.32	0.714	24.53	0.765
House	25.57	0.780	32.30	0.872	33.24	0.884	31.26	0.861	31.04	0.852	29.63	0.858	32.10	0.888
Parrots	25.42	0.832	28.96	0.900	29.67	0.911	29.43	0.908	29.33	0.904	28.67	0.907	30.50	0.928
Bike	20.68	0.613	23.98	0.770	24.64	0.796	24.04	0.786	24.26	0.783	23.83	0.794	25.33	0.843
Flower	24.37	0.687	28.48	0.829	29.09	0.848	28.80	0.844	28.72	0.836	28.05	0.843	29.85	0.883
Girl	29.72	0.738	33.26	0.819	33.54	0.827	33.43	0.825	33.27	0.819	32.66	0.830	34.34	0.861
Hat	27.04	0.802	30.57	0.872	31.41	0.886	30.62	0.872	30.71	0.873	30.05	0.879	31.56	0.902
Parents	25.05	0.789	29.90	0.890	30.65	0.902	30.08	0.898	30.13	0.895	29.08	0.899	31.44	0.928
Raccoon	26.44	0.649	28.89	0.755	29.22	0.766	29.05	0.764	28.97	0.760	29.20	0.804	30.59	0.850
Average	24.57	0.709	28.42	0.815	29.00	0.830	28.34	0.821	28.31	0.816	27.88	0.836	29.60	0.871

the image *Bike* reconstructed by GR and CNN methods, as shown in Figs. 5(c) and 6(c). We see that the reconstructed HR images by the ANR method [31] seems to be blurry. Meanwhile, the ASDS method [23] is effective in suppressing the ringing artifacts, and the edges are clearer than the other compared methods except the proposed. Compared with the other example learning-based SR methods, the proposed can effectively recover the high-frequency details lost in LR images by learning the corresponding relations between LR and HR images, as well as the spatial structure relations in HR images, showing clearer results.

Next, we discuss the images recovered under 3 \times magnification. As shown in Fig. 7(c) and (f), the HR *Fence* image reconstructed by CNN method shows many jaggy and ringing artifacts. Although the ANR method is effective to suppress the ringing artifacts, but the reconstructed results appears to be blurry, as shown in Figs. 7(d) and 8(d). The ASDS method produces very sharp edges, and makes the reconstructed image look natural. As shown in Fig. 7, the

fence is located at the border of the chimney, and the tree is restored only by SORM, while other methods cannot do this. This illustrates the effectiveness of our method in terms of remitting aliasing problem. Looking at the texture area below the parrot eye in Fig. 8, the result of A+ looks good, but the discontinuity part of the black stripe is not restored, and the jagged black hair is smoothed. Only our method restored these details even without the NL part, while ASDS that also uses the NL part cannot do this. From Figs. 7 and 8, our results show both sharper edges and finer details compared with other methods. The results of our proposed method have minimal block artifacts and show the pleasing visual quality, which is mainly because our SORM model is capable of preserving the integrality of HR pixels in one patch and discovering the inherent relatedness between pixels.

D. Statistical Experiments

In this section, we conduct a statistical experiment on 200 HR images selected from BSDS500 image database.

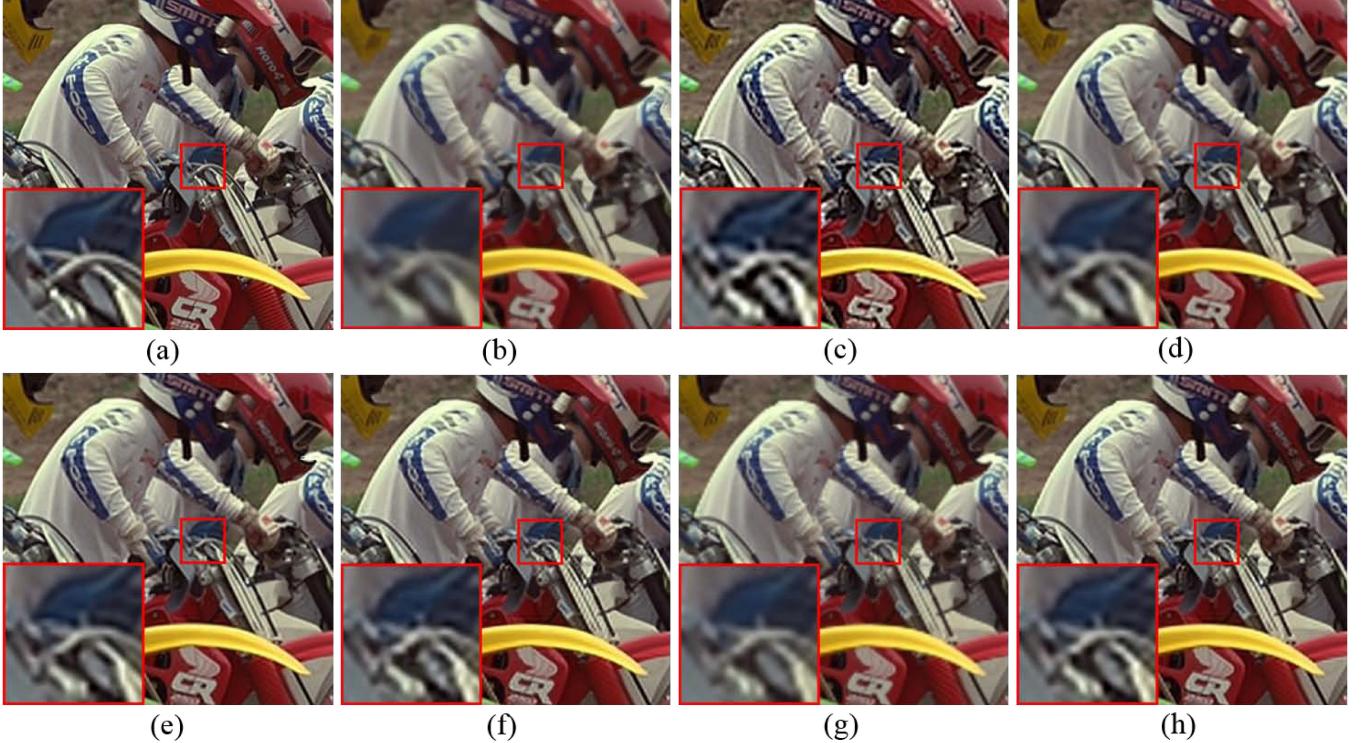


Fig. 5. Reconstructed HR images of *Bike* by different methods ($2\times$). (a) Original. (b) GPR. (c) GR. (d) ANR. (e) ASDS. (f) CNN. (g) SORM. (h) SORM + NL.

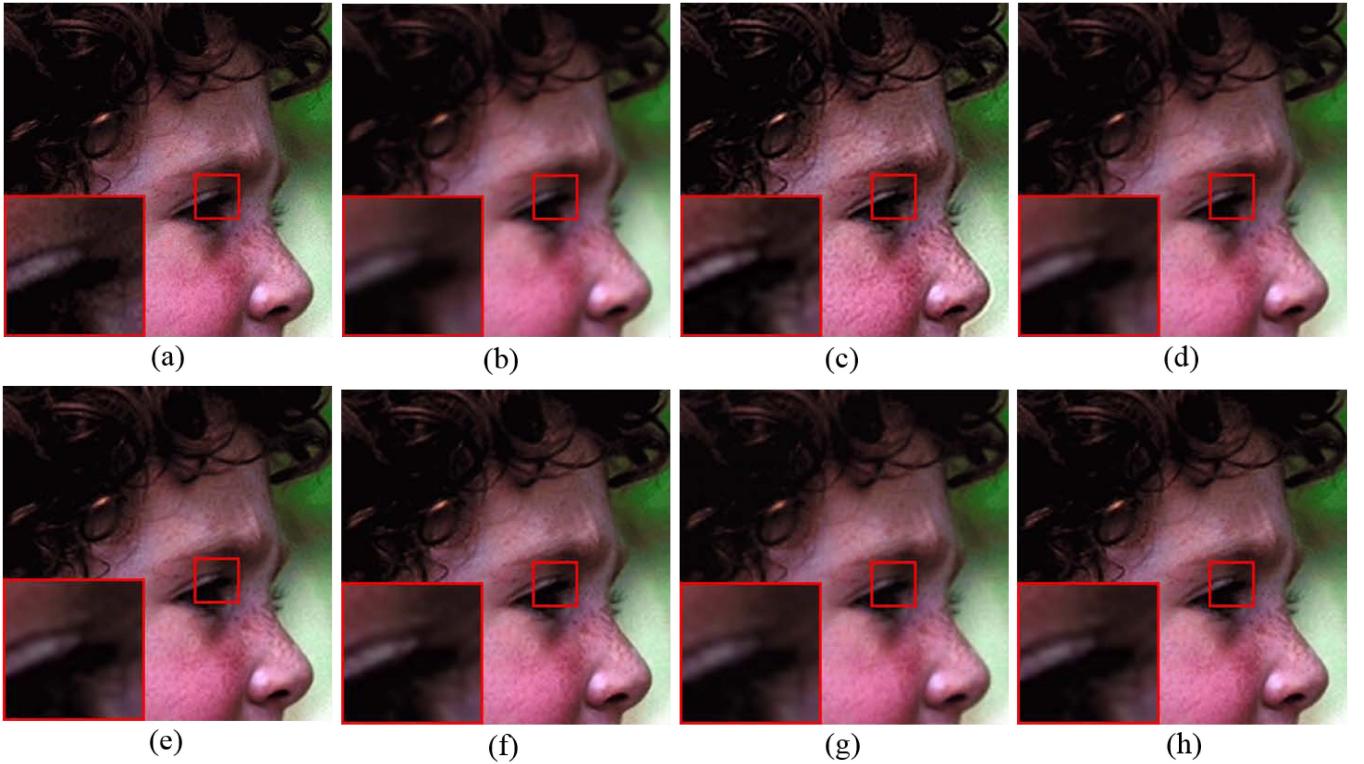


Fig. 6. Reconstructed HR images of *Girl* by different methods ($2\times$). (a) Original. (b) GPR. (c) GR. (d) ANR. (e) ASDS. (f) CNN. (g) SORM. (h) SORM + NL.

We still choose the above five representative methods as comparison baselines. We use the probability distributions of PSNR and SSIM gains to verify the effectiveness of

the proposed method. In particular, we first compute the PSNR and SSIM gains of 200 test images over comparison methods. Then, the ranges of both indices are evenly divided

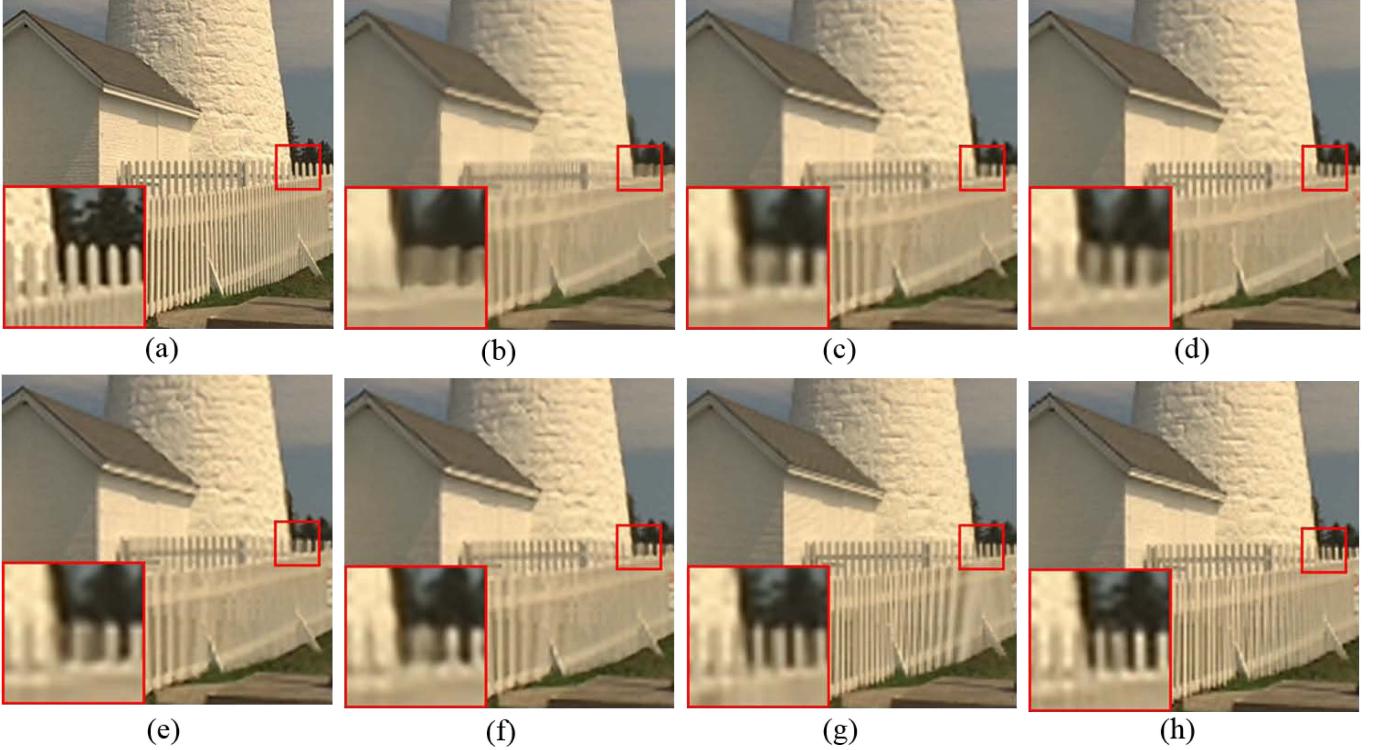


Fig. 7. Reconstructed HR images of *Fence* by different methods ($3\times$). (a) Original. (b) GPR. (c) ANR. (d) A+. (e) ASDS. (f) CNN. (g) SORM. (h) SORM + NL.

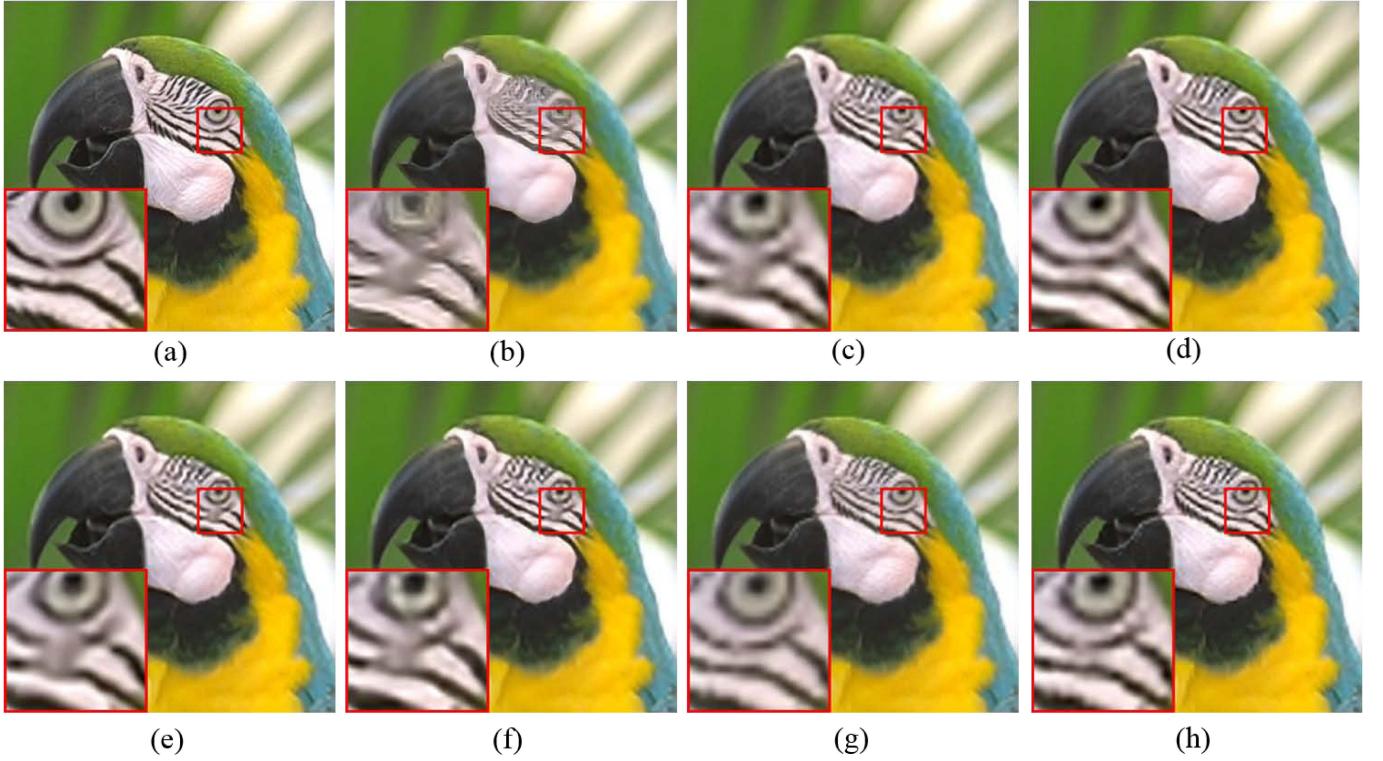


Fig. 8. Reconstructed HR images of *Parrots* by different methods ($3\times$). (a) Original. (b) GPR. (c) ANR. (d) A+. (e) ASDS. (f) CNN. (g) SORM. (h) SORM + NL.

into 20 bins. Finally, the histograms of all the bins are calculated and drawn for comparison.

As shown in Fig. 9, both the PSNR and SSIM gains over other compared methods are significantly outstanding

in all cases, especially superior to GPR. Concretely, the average statistics of the PSNR and SSIM gains over GPR are 4.36 and 0.167. The PSNR and SSIM gains over GR are 3.44 and 0.086, respectively. The PSNR and SSIM gains over

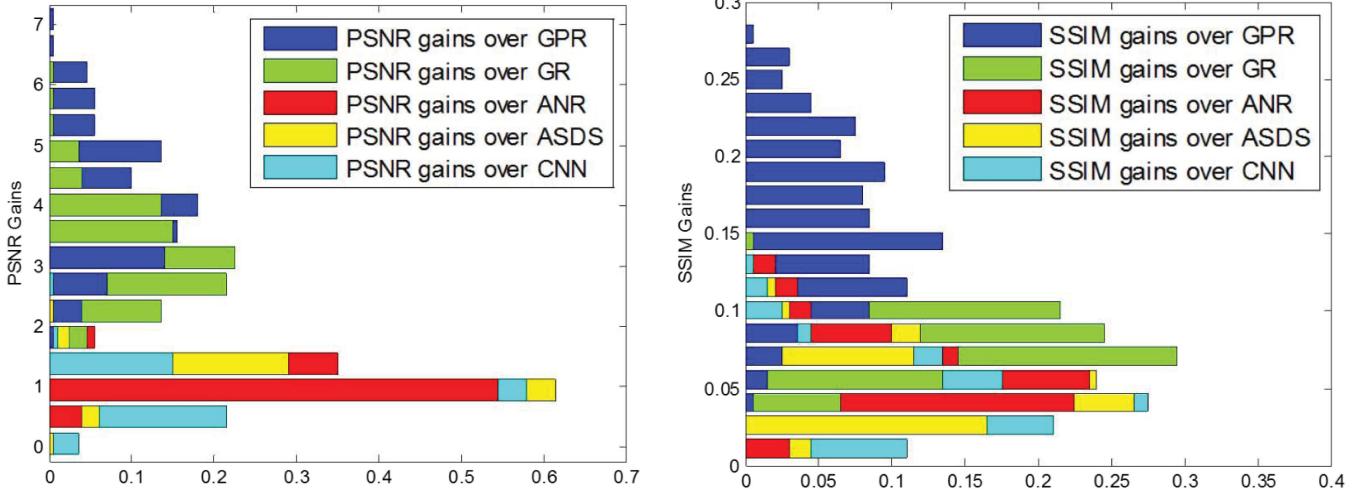


Fig. 9. Probability distributions of PSNR and SSIM gains for 200 test images in BSDS 500 database.

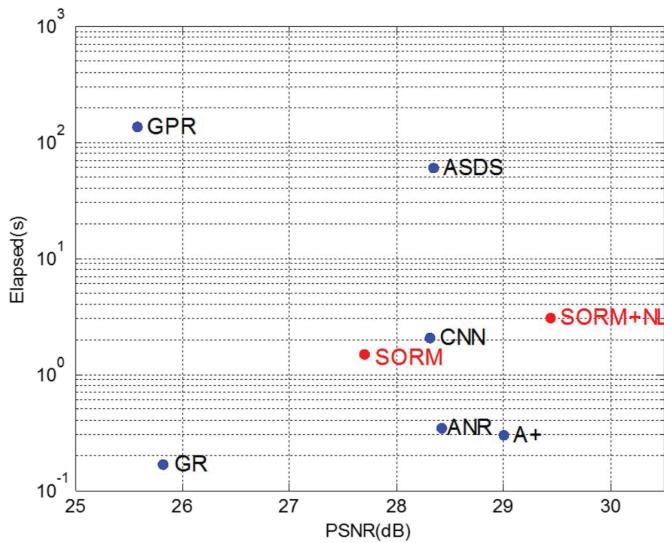


Fig. 10. Speed versus PSNR between the compared methods.

ANR are 1.34 and 0.064. The PSNR and SSIM gains over ASDS are 1.28 and 0.060. The PSNR and SSIM gains over CNN are 1.11 and 0.053. The above evaluation confirms that the proposed method is a promising SR method.

E. Computational Complexity

In this section, we compare the CPU running time and the PSNR values between our proposed method and the baselines in the SR stage. We implement our proposed method and the compared approaches on the MATLAB platform on a PC with Intel Core i5-3470 CPU of 3.2 GHz and 8-G RAM. The comparisons of $3 \times$ SR magnification on *Parrots* image are shown in Fig. 10. Because we use linear kernel function and acceleration algorithm of NL regularization term during the SR process, it takes only about 3s to recover a 256×256 HR image, which is acceptable in practical use. We can find that the proposed method costs less time than the GPR and ASDS methods and only slightly slower than

the CNN and ANR methods. It is worth noting that the reconstructed results of the proposed method are the best.

VI. CONCLUSION

In this paper, we regard the single-image SR reconstruction problem as a multioutput regression problem, which aims to learn a set of mapping functions from a multivariate input space (i.e., LR image patches) to its corresponding multivariate output space (i.e., HR image patches), simultaneously. We utilize the SORM model to build the relation between the HR and LR patches, which effectively describes the situation that a pixel of the HR patch may affect the other pixels within the same HR patch. Compared with the standard SVM, the SORM model adopts the quadratic component as the optimizing index, which greatly reduces the number of tuning parameters, and thus improves the computational efficiency as well as reconstruction quality to a great extent. In addition, considering that there are many repetitive patterns throughout a natural image, we formulate NL similarity redundancy as a regularization term to further improve the reconstruction quality with sharper edges and fine-grained details. Extensive experimental results demonstrate that the proposed SR method outperforms several state-of-the-art methods, both quantitatively and qualitatively.

Although we adopt the linear kernel function in this paper, the kernel selection is still a critical problem for regression model. In the future, we will apply the multikernel regression to learn the mapping functions, which will avoid kernel selection through a large number of cross validations. The combination of different kernels is expected to further improve the performance of kernel regression-based image SR.

REFERENCES

- [1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
- [2] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

- [3] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, no. 2, pp. 317–339, 1984.
- [4] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991.
- [5] B. C. Tom, N. P. Galatsanos, and A. K. Katsaggelos, "Reconstruction of a high resolution image from multiple low resolution images," in *Super-Resolution Imaging*. New York, NY, USA: Springer-Verlag, 2002, pp. 73–105.
- [6] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, Jun. 1996.
- [7] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. Amer. A*, vol. 6, no. 11, pp. 1715–1726, 1989.
- [8] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframe," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 6, pp. 1013–1027, Jun. 1990.
- [9] M. Unser, A. Aldroubi, and M. Eden, "Fast B-spline transforms for continuous image representation and interpolation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 3, pp. 277–285, Mar. 1991.
- [10] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin, "Super resolution using edge prior and single image detail synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2400–2407.
- [11] 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.
- [12] A. Marquina and S. J. Osher, "Image super-resolution by TV-regularization and Bregman iteration," *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, Dec. 2008.
- [13] R. O. Lane, "Non-parametric Bayesian super-resolution," *IET Radar, Sonar, Navigat.*, vol. 4, no. 4, pp. 639–648, Aug. 2010.
- [14] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian image super-resolution, continued," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1089–1096.
- [15] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.
- [16] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun./Jul. 2004, pp. 1–8.
- [17] D.-H. Trinh, M. Luong, F. Dibos, J.-M. Rocchisani, C.-D. Pham, and T. Q. Nguyen, "Novel example-based method for super-resolution and denoising of medical images," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1882–1895, Apr. 2014.
- [18] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [19] 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.
- [20] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. 7th Int. Conf. Curves Surf.*, 2010, pp. 711–730.
- [21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [22] D. Datsenko and M. Elad, "Example-based single document image super-resolution: A global MAP approach with outlier rejection," *Multidimensional Syst. Signal Process.*, vol. 18, nos. 2–3, pp. 103–121, Sep. 2007.
- [23] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.
- [24] J. Yu, X. Gao, D. Tao, X. Li, and K. Zhang, "A unified learning framework for single image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 780–792, Apr. 2014.
- [25] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with multiscale similarity learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1648–1659, Oct. 2013.
- [26] K. I. Kim and Y. Kwon, "Example-based learning for single-image super-resolution," in *Pattern Recognition*. Berlin, Germany: Springer-Verlag, Jun. 2008, pp. 456–465.
- [27] K. S. Ni and T. Q. Nguyen, "Image superresolution using support vector regression," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1596–1610, Jun. 2007.
- [28] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.
- [29] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.
- [30] H. He and W.-C. Siu, "Single image super-resolution using Gaussian process regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 449–456.
- [31] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.
- [32] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. 12th Asian Conf. Comput. Vis.*, Nov. 2014, pp. 111–126.
- [33] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. 13th Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 184–199.
- [34] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.
- [35] N. Efrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin, "Accurate blur models vs. image priors in single image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2832–2839.
- [36] J. Lee, B. Choi, and E. Kim, "Novel range-free localization based on multidimensional support vector regression trained in the primal space," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1099–1113, Jul. 2013.
- [37] P. Bouboulis, S. Theodoridis, C. Mavroforakis, and L. Evaggelatou-Dalla, "Complex support vector machines for regression and quaternary classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1260–1274, Jun. 2014.
- [38] M. Brudnak, "Vector-valued support vector regression," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2006, pp. 1562–1569.
- [39] E. Vazquez and E. Walter, "Multi-output support vector regression," in *Proc. 13th IFAC Symp. Syst. Identificat.*, Aug. 2003, pp. 1820–1825.
- [40] F. Pérez-Cruz, G. Camps-Valls, E. Soria-Olivas, J. J. Pérez-Ruixo, A. R. Figueiras-Vidal, and A. Artés-Rodríguez, "Multi-dimensional function approximation and regression estimation," in *Proc. 12th Int. Conf. Artif. Neural Netw.*, Dec. 2002, pp. 757–762.
- [41] M. Sánchez-Fernández, M. de Prado-Cumplido, J. Arenas-García, and F. Pérez-Cruz, "SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2298–2307, Aug. 2004.
- [42] A. Vedaldi, M. Blaschko, and A. Zisserman, "Learning equivariant structured output SVM regressors," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 959–966.
- [43] X. An, S. Xu, L.-D. Zhang, and S.-G. Su, "Multiple dependent variables LS-SVM regression algorithm and its application in NIR spectral quantitative analysis," *Spectrosc. Spectral Anal.*, vol. 29, no. 1, pp. 127–130, Jan. 2009.
- [44] S. Xu, X. An, X. Qiao, L. Zhu, and L. Li, "Multi-output least-squares support vector regression machines," *Pattern Recognit. Lett.*, vol. 34, no. 9, pp. 1078–1084, Jul. 2013.
- [45] S. Xu, X. An, X. Qiao, and L. Zhu, "Multi-task least-squares support vector machines," *Multimedia Tools Appl.*, vol. 71, no. 2, pp. 699–715, Jul. 2014.
- [46] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
- [47] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2272–2279.
- [48] W. Dong, L. Zhang, G. Shi, and X. Wu, "Nonlocal back-projection for adaptive image enlargement," in *Proc. 16th IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 349–352.
- [49] J. Darbon, A. Cunha, T. F. Chan, S. Osher, and G. J. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," in *Proc. 5th IEEE Int. Symp. Biomed. Imag.*, May 2008, pp. 1331–1334.
- [50] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.



Cheng Deng received the B.Sc., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China.

He is currently a Full Professor with the School of Electronic Engineering, Xidian University. He has authored or co-authored over 40 scientific articles at top venues, including the IEEE TNNLS, TMM, TIP, TCYB, ICCV, CVPR, AAAI, and IJCAI. His current research interests include computer vision, multimedia processing and analysis, and information hiding.



Dacheng Tao (F'15) is Professor of Computer Science with the Centre for Quantum Computation and Intelligent Systems, and the Faculty of Engineering and Information Technology in the University of Technology, Sydney. He mainly applies statistics and mathematics to data analytics problems and his research interests spread across computer vision, data science, image processing, machine learning, and video surveillance. His research results have expounded in one monograph and 100+ publications at prestigious journals and prominent conferences,

such as IEEE T-PAMI, T-NNLS, T-IP, JMLR, IJCV, NIPS, ICML, CVPR, ICCV, ECCV, AISTATS, ICDM; and ACM SIGKDD, with several best paper awards, such as the best theory/algorithm paper runner up award in IEEE ICDM'07, the best student paper award in IEEE ICDM'13, and the 2014 ICDM 10 Year Highest-Impact Paper Award.



Jie Xu received the B.Sc. degree in biomedical engineering and the M.Sc. degree in circuits and systems from Xidian University, Xi'an, China, in 2012 and 2015, respectively, where she is currently pursuing the Ph.D. degree with the School of Electronic Engineering.

Her current research interests include pattern recognition and computer vision.



Xinbo Gao (M'07–SM'12) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

He was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan, from 1997 to 1998. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is currently a Cheung Kong Professor with the Ministry of Education, Beijing, China, a Professor of Pattern Recognition and Intelligent System, and the Director of the State Key Laboratory of Integrated Services Networks. He has authored five books and around 200 technical articles in his research areas in refereed journals and proceedings, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, the *International Journal of Computer Vision*, and *Pattern Recognition*. His current research interests include computational intelligence, machine learning, computer vision, pattern recognition, and wireless communications.

Prof. Gao is a fellow of the Institution of Engineering and Technology. He is on the editorial boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He served as the General Chair/Co-Chair or Program Committee Chair/Co-Chair or PC Member for around 30 major international conferences.



Kaibing Zhang received the M.Sc. degree in computer software and theory from Xihua University, Chengdu, China, in 2005, and the Ph.D. degree from Xidian University, Xi'an, China, in 2012.

He has been a Post-Doctoral Research Fellow with the School of Electronic Engineering, Xidian University, since 2013. He is currently an Associate Professor with the School of Computer and Information Science, Hubei Engineering University, Xiaogan, China. His current research interests include pattern recognition and computer vision.

Xuelong Li (M'02–SM'07–F'12) is currently a Full Professor with the Center for Optical Imagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi, China.