Single Image Super-Resolution via Locally Regularized Anchored Neighborhood Regression and Nonlocal Means

Junjun Jiang, *Member, IEEE*, Xiang Ma, Chen Chen, Tao Lu, Zhongyuan Wang, *Member, IEEE*, and Jiayi Ma, *Member, IEEE*

***Abstract*—The goal of learning-based image super resolution (SR) is to generate a plausible and visually pleasing high-resolution (HR) image from a given low-resolution (LR) input. The SR problem is severely underconstrained, and it has to rely on examples or some strong image priors to reconstruct the missing HR image details. This paper addresses the problem of learning the mapping functions (i.e., projection matrices) between the LR and HR images based on a dictionary of LR and HR examples. Encouraged by recent developments in image prior modeling, where the state-of-the-art algorithms are formed with nonlocal self-similarity and local geometry priors, we seek an SR algorithm of similar nature that will incorporate these two priors into the learning from LR space to HR space. The nonlocal self-similarity prior takes advantage of the redundancy of similar patches in natural images, while the local geometry prior of the data space can be used to regularize the modeling of the nonlinear relationship between LR and HR spaces. Based on the above two considerations, we first apply the local geometry prior to regularize the patch representation, and then utilize the nonlocal means filter to improve the super-resolved outcome. Experimental results verify the effectiveness of the proposed algorithm compared with the state-of-the-art SR methods.**

***Index Terms*—Anchored neighborhood regression, locality geometry, neighbor embedding, nonlocal means, super-resolution (SR).**

Manuscript received August 31, 2015; revised May 29, 2016 and July 21, 2016; accepted July 23, 2016. Date of publication August 10, 2016; date of current version December 14, 2016. This work was supported by the National Natural Science Foundation of China under Grant 61501413, Grant 61502354, Grant 61671332, and Grant 61503288, by the China Fundamental Research Funds for the Central Universities under Grant 310824153508, by the Shannxi Science Foundation of China under Grant 2015JM6309, and by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUGL160412. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shahram Shirani. *(Corresponding author: Xiang Ma.)*

J. Jiang is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China, and also with the Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China (e-mail: junjun0595@163.com).

X. Ma is with the School of Information Engineering, Chang'an University, Xi'an 710048, China (e-mail: maxiangmail@163.com).

C. Chen is with the Center for Research in Computer Vision, University of Central Florida, Orlando, FL 32816 USA (e-mail:chenchen870713@gmail.com).

T. Lu is with the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430073, China (e-mail: lutxyl@gmail.com).

Z. Wang is with the National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan 430072, China (e-mail: wzy hope@163.com).

J. Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jyma2010@gmail.com).

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/TMM.2016.2599145

I. Introduction

IMAGES with high-resolution (HR) are desired and often required. They can offer more details that may be critical in various applications, such as remote sensing [1], [2], medical diagnostic [3], intelligent surveillance [4]–[7]. Given an observed LR image, how to induce an HR image is an active research topic in the image processing community. Super-resolution (SR) reconstruction is a technology, which was first proposed by Huang *et al.* [8] in 1984, can estimate HR images from observed LR images. It increases the high frequency components and removes the degradations caused by the imaging process of camera sensors.

A. State of Research

Generally speaking, existing image SR methods can be classified into three categories [9]: interpolation-based SR methods, reconstruction-based multi-image SR methods, and learning-based single image SR (SISR) methods.

To super-resolve an LR observation, interpolation-based SR (e.g., Bilinear, Bicubic, and other resampling methods) utilizes a base function or an interpolation kernel to estimate the unknown pixels in HR grids. Although these approaches are very simple and fast, they are prone to blur high-frequency details and therefore may lead to noticeable blurring edges and unclear textures in the super-resolved HR image. Recently, a wide range of approaches that outperform functional interpolation have been developed, such as geometric regularity of image structures [10], [11] and gradient profile priors [12], [13].

In order to introduce more prior knowledge, reconstruction-based multi-image SR methods combine the non-redundant information contained in multiple LR images to generate an HR one. Due to the fact that image degeneration process has information loss and numerous pixel intensities need to be predicted from the limited input data, reconstruction-based multi-image SR is also a challenging and severely ill-posed problem [14], [15]. Recently, many prior knowledge has been introduced to facilitate the reconstruction process to enhance the SR performance, such as the projection-onto-convex-sets (POCS) approach [16], iterative back projection (IBP) approach [17], and adaptive filtering approach [18]. However, when the motion between two input LR images is estimated inaccurately, which often occurs in non-global motion fields [19]–[21], annoying artifacts will appear in the super-resolved results. With the increasing of magnification factor, the reconstruction constraints

and smoothness prior will provide less useful information. They tend to smear out image details and sometimes have difficulty with recovering fine structures and textures [22]. It is also stated that for large enough magnification factors, any smoothness prior will lead to overly smooth results with very little high-frequency content. As reported in [23], the practical limit of magnification factor for reconstruction-based multi-image SR methods is 1.6 if the noise removal and image registration is not good enough in the pre-processing.

Learning-based or example-based SISR methods assume that high-frequency details lost in the input LR image can be learnt from a training set of LR and HR image pairs. According to the differences of learning strategies, they can be divided into two categories: explicitly regression based [24]–[26] and implicitly coding based [5], [27], [28]. The explicit regression relationship or the implicit coding relationship between LR images and the corresponding HR ones can be used to predict the missing HR frequencies in the LR observation. In recent years, many different prior constraints have been introduced to regularize the under-constrained and ill-posed explicit regression learning or implicit coding. Priors that are commonly exploited in SISR methods mainly include gradient profile prior [29], soft information regularization [30] and sparsity either through Gaussian mixture models [31] or through an analysis operator [32]. Liu *et al.* [33], [34] took advantage of contextual information of local patches based on a Markov random fields (MRFs) model, along with structure-modulated image priors. Dong *et al.* [35], [36] combined the ideas of data clustering, adaptive principal components analysis (PCA) based sparse representations and nonlocal self-similarity of image patches within a given image.

In machine learning and pattern recognition, locality prior of training data is critical for exploring the geometry of data. In particular, the locality constraint can help reveal the nonlinear manifold structure of the data space [37]–[39]. The locality-constrained algorithms try to embed the relationship in high dimensional space to low dimensional space, making that nearby points in high dimensional space remain nearby and similarly co-located with respect to one another in low dimensional space [40], [41]. This is the essential of image SR reconstruction, which states that the high-frequency details lost in an LR image can be learnt from a training set of LR and HR image pairs, i.e., the relationship between training LR image patches and the corresponding HR patches can be used to estimate the missing HR components in the given LR input [42].

Another property of natural images is the nonlocal correlations, i.e., self-similarity of local patch patterns. Unlike "local mean" filters, which take the mean value of a group of pixels surrounding a target pixel to smooth the image, nonlocal means (NLM) filtering takes a mean of all pixels in the image, weighted by how similar these pixels are to the target pixel. It has been successfully used for image denoising [43], inpainting [44]–[46], and image restoration [47]–[50].

B. Motivation and Contributions

In this paper, we focus on the learning-based SISR. It super-resolves the input LR image patch individually by learning the prior knowledge from the LR and HR patch pairs. We follow these works [24], [51] and learn the prior information by regression functions. Specially, in this paper we take the anchored neighborhood regression (ANR) based SR method [51] as a starting point to propose a novel SISR method by regression functions, namely locally regularized anchored neighborhood regression based SR with NLM (we call it LANR-NLM for short). Specifically, we take the nonlocal redundancies and the local geometry structure of the training data into account and develop two assembled priors to regularize the ill-posed SR reconstruction problem. By introducing the nonlocal redundancies, more robust SR estimation can be expected. Moreover, the proposed method utilizes the locality-constrained regression in place of the ridge regression in ANR. It can well explore the nonlinear relationship between the LR and HR spaces and make the regression solution more stable. The experimental results show improvement of our proposed method over the original ANR approach, for example, 0.1-0.4 dB higher than the ANR method in terms of PSNR.

The proposed LANR-NLM based SR method consists of a learning stage and a reconstruction stage as shown in Fig. 1. In the learning stage, we first collect a training set consisting of a large number of co-occurrence LR and HR images, and then extract pairs of matching patches to form the LR and HR training sets. A compact LR dictionary is trained by sparse coding, and the corresponding HR dictionary is constructed under the assumption that LR and HR features share the same representations. LANR model is then used to learn the projection matrix for each dictionary atom. In the reconstruction stage, we first search the nearest atom in the LR dictionary for each patch, and then use the learned projection matrix of this atom to predict the HR features. Finally, we employ an NLM filter enhancement approach to reduce artifacts in the estimated HR images.

Built upon our preliminary work published in [52], here we give a detailed description and evaluation of our proposed LANR-NLM method in the following aspects: (i) the introduction section is rewritten to provide an extensive review of relevant work and to make our contributions clear; (ii) the NLM filter is incorporated to the regression model to improve the performance of our original model; (iii) extensive experimental evaluations are carried out to verify the effectiveness of our proposed method compared with existing state-of-the-art SISR approaches.

C. Organization of this Paper

The remainder of the paper is organized as follows. Section II provides problem definitions and related background. Section III describes the details of our proposed LANR-NLM method. Section IV presents the experimental results and analysis. Finally, we conclude this paper in Section V.

II. Problem Definitions and Related Background

A. Image Degradation Model

To comprehensively analyze the image SR reconstruction problem, the first step is to formulate an observation model that

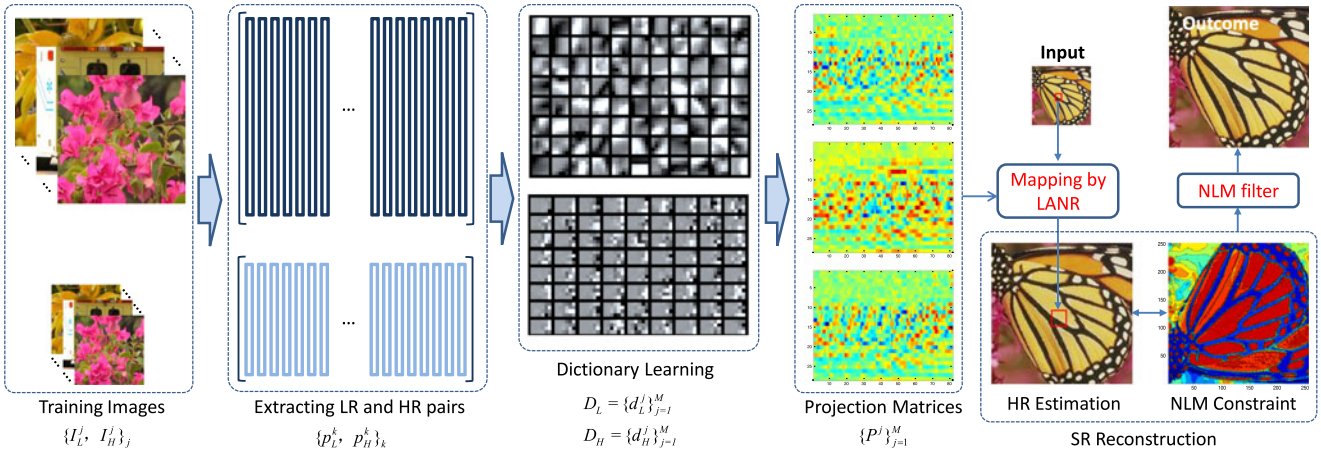


Fig. 1. Pipeline of our proposed LANR-NLM-based SR reconstruction framework.

relates the HR image to the observed LR image. Concretely, let $\mathbf{x}$ and $\mathbf{y}$ denote an HR image and the corresponding LR image, respectively. The relationship between the original HR image $\mathbf{x}$ and the LR observation $\mathbf{y}$ can be mathematically modeled by the following expression:

$$\mathbf{y} = \mathbf{DHx} + \mathbf{v} \tag{1}$$

where $\mathbf{H}$ is a blurring filter for the HR image, $\mathbf{D}$ is a matrix representing the decimation operator, and $\mathbf{v}$ is the additive Gaussian white noise accounting for the imaging sensor noise. Here, $\mathbf{x}$ and $\mathbf{y}$ are lexicographically stacked versions of the original HR image and the observed LR image, respectively.

Due to the ill-posed nature of SR reconstruction problem, the regularization-based techniques have been widely used to regularize the solution spaces. In order to obtain an effective regularizer, it is of great importance to find and model an appropriate prior knowledge of natural images, denoted by $\Omega(\mathbf{x})$. The learned prior knowledge can be employed to regularize the solution

$$J(\mathbf{x}) = ||\mathbf{y} - \mathbf{DHx}||_2^2 + \lambda\Omega(\mathbf{x}) \tag{2}$$

where λ is a Lagrangian multiplier parameter, which balances the tradeoff between the regularization term $\Omega(\mathbf{x})$ and the likelihood term $||\mathbf{y} - \mathbf{DHx}||_2^2$.

Various image prior models have been developed. Tikhonov regularization [53] and total variation [54] regularization are two popular explicit regularization techniques, which are predefined, i.e., letting $\Omega(\mathbf{x}) = ||\mathbf{Cx}||_2^2$, or $\Omega(\mathbf{x}) = ||\mathbf{Cx}||_1$, where $\mathbf{C}$ is a Laplacian operator. However, those methods based on pre-defined priors tend to smooth image details and can hardly lead to satisfactory results. This is mainly because it is difficult to know what and how much prior information should be used for the SR problem in advance [55]. Another class of prior knowledge learned from a set of LR and HR image pairs instead of using pre-defined ones usually lead to better SR reconstruction results. These techniques assume that an image can be locally or sparsely represented in some domain spanned by a set of bases [9], [27], [56]. In particular, they first learn coding coefficients in the LR space, and then apply the same coding coefficients to the HR space to obtain the target HR image. In the following, we will briefly review several representative coding based image SR approaches.

B. Coding-Based Image SR

Following the notation used in [55], for an HR image $\mathbf{x} \in R^N$, let $\mathbf{x}_i = \mathbf{R}_i\mathbf{x}$ denote an HR image patch of size $\sqrt{n} \times \sqrt{n}$ extracted at location i, where $\mathbf{R}_i$ is the matrix extracting patch $\mathbf{x}_i$ from $\mathbf{x}$ at location i. Similarly, for an LR image $\mathbf{y} \in R^{N/s^2}$, let $\mathbf{y}_i = \mathbf{R}_i\mathbf{y}$ denote an LR image patch of size $\sqrt{n}/s \times \sqrt{n}/s$ extracted at location i. Here, s is the down-sampling factor.

Given the LR and HR dictionaries $\mathbf{D}_L = [\mathbf{d}_L^1, \mathbf{d}_L^2, ..., \mathbf{d}_L^M]$ and $\mathbf{D}_H = [\mathbf{d}_H^1, \mathbf{d}_H^2, ..., \mathbf{d}_H^M]$, where M is the dictionary size, the key issue of coding based methods is how to represent an image patch $\mathbf{y}_i$ in the LR space and obtain the optimal weight vector $\mathbf{w}$

$$\hat{\mathbf{w}}_i = \arg\min_{\mathbf{w}_i} ||\mathbf{y}_i - \mathbf{D}_L\mathbf{w}_i||_2^2 + \lambda\Omega(\mathbf{w}_i). \tag{3}$$

The target HR patch can be obtained by $\mathbf{x}_i = \mathbf{D}_H\mathbf{w}_i$ and the final HR image $\mathbf{x}$ is calculated by averaging all the reconstructed patches with the set of coding coefficients $\mathbf{w}_i$. Mathematically, it can be written as

$$\mathbf{x} = \left(\sum_{i=1}^{N} {\mathbf{R}_i}^T\mathbf{R}_i\right)^{-1} \sum_{i=1}^{N} {\mathbf{R}_i}^T\mathbf{D}_H\mathbf{w}_i. \tag{4}$$

For the convenience of expression, we define the operator "$\circ$" as follows:

$$\mathbf{x} \approx \mathbf{D}_H \circ \mathbf{w} = \left(\sum_{i=1}^{N} {\mathbf{R}_i}^T\mathbf{R}_i\right)^{-1} \sum_{i=1}^{N} {\mathbf{R}_i}^T\mathbf{D}_H\mathbf{w}_i \tag{5}$$

where $\mathbf{w}$ denotes the concatenation of all $\mathbf{w}_i$.

In the following, we will introduce two representative local patch based coding approaches.

1) Neighbor Embedding: Based on the assumption that LR and HR image patches lie on low-dimensional nonlinear manifolds and share similar local structures, Chang *et al.* [56] proposed to allow LR input patches to be approximated by a linear combination of their nearest neighbors in the database. Firstly, for each LR patch $\mathbf{y}_i$ in the input LR image, the optimal

reconstruction weights can be obtained by minimizing the local reconstruction error

$$\hat{\mathbf{w}}_i = \arg\min_{\mathbf{w}_i} ||\mathbf{y}_i - \sum_k w_{i,k}\mathbf{D}_L^k||_2^2$$
$$\text{s.t.} \sum_k w_{i,k} = 1 \tag{6}$$

where k is the index of nearest neighbors of $\mathbf{y}_i$ in the LR dictionary. Minimizing (6) is a constrained least squares problem, which can be transform to solve a linear system equation with the "sum to one" constraint.

Recently, Bevilacqua *et al.* [57] also used neighbor embedding for SR. They assume that the local nonnegative least squares decomposition weights over a local neighborhood in the LR space are the same as those in the HR space.

2) Sparse Coding: Neighbor embedding approaches may produce undesirable results due to over- or under-fitting problem. To this end, Yang *et al.* [27] replaced the least square estimation with a sparsity constrained optimization to obtain more accurate solution. Accordingly, the LR patches are sparsely reconstructed from a learned dictionary using the following formulation:

$$\hat{\mathbf{w}}_i = \arg\min_{\mathbf{w}_i} ||\mathbf{y}_i - \sum_{m=1}^{M} w_{i,m}\mathbf{D}_L^m||_2^2 + \lambda||\mathbf{w}||_1 \tag{7}$$

where $||\bullet||_1$ denotes the ℓ^1-norm. Here, $\mathbf{D}_L$ is the learned LR dictionary, which is different from the neighbor embedding approaches. Sparse dictionaries are jointly learned for LR and HR coupled feature spaces, with the goal of having the same sparse representation for LR patches and their corresponding HR patches.

Built upon this sparse coding based framework, Zeyde *et al.* [58] used the K-SVD algorithm to learn the LR dictionary while directly used the pseudo-inverse to obtain the HR dictionary. Moreover, they performed dimensionality reduction on the patches using PCA.

III. The Proposed Method

In order to achieve a reasonable solution to the severely ill-posed SR problem, in this paper we propose to take advantage of local geometry prior of the data space and the nonlocal self-similarity prior of natural images to regularize the SR reconstruction. In the following, we first review the ANR method, and then present our proposed locally regularized anchored neighborhood regression model with NLM.

A. Anchored Neighborhood Regression (ANR)

In [51], Timofte *et al.* supported the use of sparse learned dictionaries in combination with neighbor embedding methods, and proposed an ANR based SR approach.

In ANR, each atom of the learned dictionary (by the method of [58]) is considered as one anchor point in the LR patch space. Each anchor point is associated with a mapping function that is learned off-line. Instead of considering the whole dictionary like the sparse encoding approach in Section II-B.2, ANR proposes to work in the local neighborhood $\mathbf{N}_{i,j}^L$ of the LR dictionary

$$\mathbf{N}_{i,j}^L = \{\mathbf{D}_L^k\}_{k\in C_K(\mathbf{D}_L^j)} \tag{8}$$

where $\mathbf{D}_L^j$ denotes the nearest neighbor of the input LR patch $\mathbf{y}_i$ in the LR dictionary $\mathbf{D}_L$, and $C_K(\mathbf{D}_L^j)$ is the index set of the K nearest neighbors of $\mathbf{D}_L^j$ in the LR dictionary $\mathbf{D}_L$.

In particular, an input LR image patch is constructed as a least squares regression problem regularized by the l_2-norm of the reconstruction vector. Thus

$$\hat{\mathbf{w}}_i = \arg\min_{\mathbf{w}_i} ||\mathbf{y}_i - \mathbf{N}_{i,j}^L\mathbf{w}_i||_2^2 + \lambda_1||\mathbf{w}||_2^2. \tag{9}$$

Here, λ_1 is a regularization parameter, which balances the trade-off between the reconstruction error of $\mathbf{y}_i$ and the smoothness of $\hat{\mathbf{w}}_i$.

Equation (9) is a ridge regression problem and it has a closed-form solution

$$\mathbf{w_i} = (\mathbf{N_{i,j}^L}^{\boldsymbol{T}}\mathbf{N_{i,j}^L} + \lambda_1\mathbf{I})^{-1}\mathbf{N_{i,j}^L}^{\mathbf{T}}\mathbf{y_i}. \tag{10}$$

The HR patches can be computed using the same reconstruction weights on the HR neighborhood $\mathbf{N}_j^H$

$$\mathbf{x_i} = \mathbf{N_j^H}\mathbf{w_i}, \tag{11}$$

where $\mathbf{x}_i$ is the HR output patch and $\mathbf{N}_j^H$ is the HR neighborhood corresponding to $\mathbf{N}_{i,j}^L$.

From (10) and (11), we obtain

$$\mathbf{x}_i = \mathbf{N_j^H}(\mathbf{N}_{i,j}^{L\,T}\mathbf{N}_{i,j}^L + \lambda_1\mathbf{I})^{-1}\mathbf{N}_{i,j}^{L\,T}\mathbf{y}_i. \tag{12}$$

Let $\mathbf{P}^j = \mathbf{N_j^H}(\mathbf{N}_{i,j}^{L\,T}\mathbf{N}_{i,j}^L + \lambda_1\mathbf{I})^{-1}\mathbf{N}_{i,j}^{L\,T}$ be the projection matrix for dictionary atom $\mathbf{D}_L^j$, then the coding based SR problem is transformed to the regression problem. Therefore, we can calculate the projection matrix $\mathbf{P}^j$ for each dictionary atom $\mathbf{D}_L^j$, $j = 1, 2, ..., M$, based on its own neighborhood $\mathbf{N}_{i,j}^L$.

Upon acquiring all the projection matrices, the SR problem can be solved by mapping the input LR patch into the HR space

$$\mathbf{x}_i = \mathbf{P}^j\mathbf{y}_i. \tag{13}$$

B. Locally Regularized Anchored Neighborhood Regression for Regularization

From (9), we learn that ANR treats all the neighbors in $\mathbf{N}_{i,j}^L$ equally. Thus, it is not flexible and adaptive to the input patch when obtaining the mapping function. To obtain more accurate reconstruction weights, in this paper we introduce more prior information to the ridge regression problem (9). Locality has been verified to be a very important property for exploring the nonlinear data structure. In fact, [4], [37], [38] have shown that locality is more essential than sparsity, as locality must lead to sparsity but not necessary vice versa.

Inspired by this, we introduce the locality regularization to the objective function

$$\hat{\mathbf{w}}_i = \arg\min_{\mathbf{w}_i} ||\mathbf{y}_i - \mathbf{N}_{i,j}^L\mathbf{w}_i||_2^2 + \lambda_1||\mathbf{g}_i \odot \mathbf{w}_i||_2^2$$
$$\text{s.t.}\ \mathbf{1}^T\mathbf{w}_i = 1 \tag{14}$$

Algorithm 1: Projection Matrix Learning.

Input:
A set of HR training images $\{\mathbf{I}_H^j\}_j$.
Output:
Projection matrices $\{\mathbf{P}^j\}_j$.
1: Construct LR images, $\{\mathbf{I}_L^j\}_j$, by a 7×7 Gaussian blurring operator with a standard deviation of 1.6 and decimated by a factor of 3.
2: Extract pairs of matching patches form the LR and HR training databases, $\{\mathbf{p}_H^k, \mathbf{p}_L^k\}_k$.
3: Remove low-frequencies from $\mathbf{p}_H^k$ and extract the high frequency information of LR patches $\mathbf{p}_L^k$.
4: Project the extracted features onto a low-dimensional subspace using PCA dimensionality reduction.
5: Train a dictionary $\mathbf{D}_L$ for the LR patches, such that they can be represented sparsely.
6: Construct the corresponding HR dictionary $\mathbf{D}_H$ for the LR patches, such that it matches the LR one.
7: **for** each LR atom $\mathbf{D}_L^j$ of $\mathbf{D}_L$ **do**
8: Search $\mathbf{D}_L^j$ over $\mathbf{D}_L$ to obtain the K LR nearest neighbors and the penalize vector: $\mathbf{N}_{i,j}^L = \{\mathbf{D}_L^k\}_{k \in C_K(\mathbf{D}_L^j)}, g_{j,k} = \{1/corr(\mathbf{D}_L^j, \mathbf{D}_L^k)\}^\alpha$, $k \in C_K(\mathbf{D}_L^j)$.
9: Obtain the corresponding K LR nearest neighbors in HR dictionary $\mathbf{D}_H$: $\mathbf{N}_j^H = \{\mathbf{D}_H^k\}_{k \in C_K(\mathbf{D}_L^j)}$.
10: Calculate $\mathbf{g}_j = diag(g_{j,1}, g_{j,2}, ..., g_{j,K})$.
11: Calculate the projection matrix: $P^j = \mathbf{N}_{\boldsymbol{j}}^{\boldsymbol{H}} ({\mathbf{N}_{i,j}^L}^T \mathbf{N}_{i,j}^L + \lambda \mathbf{g}_j)^{-1} {\mathbf{N}_{i,j}^L}^T$.
12: **end for**
13: Output the projection matrices $\{\mathbf{P}^j\}_j$.

Algorithm 2: Image SR via LANR-NLM.

Input:
LR and HR dictionaries, $\mathbf{D}_L$ and $\mathbf{D}_H$ [1] and an LR test image y. The regularization parameters λ_1, λ_2, α and nearest neighbor number K, $maxIter$, e.
Output:
HR target image $\mathbf{x}$.
1: **for** each LR patch $\mathbf{y}_i$ of $\mathbf{y}$ **do**
2: Search $\mathbf{y}_i$ over $\mathbf{D}_L$ to find the nearest neighbor and its position j (i.e., the subscript of the nearest neighbor).
3: Compute the HR version of $\mathbf{y}_i$ via linear mapping, $\mathbf{x}_i = \mathbf{P}_j \mathbf{y}_i$.
4: **end for**
5: Integrate all the reconstructed HR patches $\mathbf{x}_i$ and average pixel values in the overlap regions to form the HR image $\mathbf{x}$.
6: Adopt the gradient descent rule to refine the output of Step 5 to get the target HR version $\mathbf{x}$.
7: **repeat**
8: $\hat{\mathbf{x}}^{(t+1/2)} = \mathbf{x}^{(t)} + \delta \mathbf{H}^T \mathbf{D}^T (\mathbf{y} - \mathbf{D}\mathbf{H}\mathbf{x}^{(t)})$, where δ is the pre-determined constant.
9: *if* mod(t, M_0)
10: Update the matrix B using the improved estimation $\hat{\mathbf{x}}^{(t+1/2)}$.
11: *end*
12: $\hat{\mathbf{x}}^{(t+1)} = (\mathbf{I} - \mathbf{B})\mathbf{x}^{(t+1/2)}$.
13: $t = t + 1$.
14: **until** $t > maxIter$ or $||\mathbf{x}^t - \mathbf{x}^{t+1}||_2^2 / N < e$
15: Output the reconstructed HR image $\hat{\mathbf{x}}$

where "$\odot$" denotes a point wise vector product, λ_1 is a parameter to balance the tradeoff between the reconstruction error of $\mathbf{y}_i$ and the locality of the solution of $\hat{\mathbf{w}}_i$. The constraint $\mathbf{1}^T \mathbf{w}_i = 1$ follows the shift-invariant requirement, and $\mathbf{g}_i$ is a K-dimensional vector that penalizes the distance between the input LR patch $\mathbf{y}_i$ and each K nearest dictionary atom, thus giving different freedom to each dictionary atom, which is inversely proportional to the correlation $corr(\mathbf{y}_i, \mathbf{D}_L^k)$ to the input LR patch $\mathbf{y}_i$ (Note that in this paper we follow [51] to use the correlation rather than the Euclidean distance to measure the similarity between patches). Specifically

$$g_{i,k} = \{1/corr(\mathbf{y}_i, \mathbf{D}_L^k)\}^\alpha, k \in C_K(\mathbf{D}_L^j) \tag{15}$$

where α is used for adjusting the weight decay speed for the locality adaptor, which is set to 11 in all our experiments. More details about the setting of α, please refer to Section IV-B.

For all $\mathbf{y}_i$, we can rewrite (14) as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{N}^L \mathbf{w}||_2^2 + \lambda_1 ||\mathbf{G}\mathbf{w}||_2^2$$
$$\text{s.t.} \mathbf{1}^T \mathbf{w}_i = 1. \tag{16}$$

Here, $\mathbf{N}^L$ and $\mathbf{G}$ are block diagonal matrices, $\mathbf{N}^L = blkdiag(\mathbf{N}_{1,j}^L, \mathbf{N}_{2,j}^L, ..., \mathbf{N}_{N,j}^L)$ and $\mathbf{G} = blkdiag(\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_N)$.

From (1), we can easily deduce $\mathbf{y} = \mathbf{D}_L \hat{\mathbf{w}} = \mathbf{D}\mathbf{H}\mathbf{D}_H \mathbf{w}$. Similarly, we have $\mathbf{N}^L \mathbf{w} = \mathbf{D}\mathbf{H}\mathbf{N}^H \mathbf{w}$. Thus, (16) can be rewritten as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{D}\mathbf{H}\mathbf{N}^H \mathbf{w}||_2^2 + \lambda_1 ||\mathbf{G}\mathbf{w}||_2^2$$
$$\text{s.t.} \mathbf{1}^T \mathbf{w}_i = 1. \tag{17}$$

By introducing the locality prior, our proposed method can achieve more appropriate patch representation. It can obtain relative smooth reconstruction weights. In other words, large reconstruction weights will be assigned to the training patches that are similar to the input patch while small reconstruction weights will be assigned to the training patches that are dissimilar to the input patch.

C. Adaptive Regularization by Nonlocal Similarity

The LANR model can exploit the local geometry in data space. In addition, there are often many repetitive patterns throughout a natural image. Such nonlocal redundancy is very helpful to improve the quality of reconstructed images.

Therefore, we further incorporate NLM to the LANR model as a complementary regularization term.

Alternatively, for each local patch $\mathbf{x}_i$, we search for its L similar patches $\{\mathbf{x}_i^l\}_{l=1}^L$ in the whole image $\mathbf{x}$ (in practice, a large enough area around $\mathbf{x}_i$ can be used for the sake of efficiency). Then we use the obtained L similar patches to predict the patch $\mathbf{x}_i$ by

$$\mathbf{x}_i = \sum\nolimits_{l=1}^{L} \mathbf{x}_i^l b_i^l. \tag{18}$$

We set the nonlocal weight b_i^l to be inversely proportional to the distance between patches $\mathbf{x}_i$ and $\mathbf{x}_i^l$

$$b_i^l = \frac{1}{\gamma} \exp(-||\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_i^l||_2^2/h) \tag{19}$$

where $\hat{\mathbf{x}}_i = \mathbf{D}_H \hat{\mathbf{w}}_i$ and $\hat{\mathbf{x}}_i^l = \mathbf{D}_H \hat{w}_{i,l}$ are the estimates of the patches $\mathbf{x}_i$ and $\mathbf{x}_i^l$, h is a pre-determined control factor of the weight, B is the normalization factor, and $\gamma = \sum_{l=1}^{L} \exp(-||\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_i^l||_2^2/h)$.

Let $\mathbf{b}_i$ be the column vector containing all the weights b_i^l and $\boldsymbol{\beta}_i$ be the column vector containing all $\mathbf{x}_i^l$. Equation (18) can be rewritten as

$$\mathbf{x}_i = \mathbf{b}_i^T \boldsymbol{\beta}_i. \tag{20}$$

D. Summary of the Algorithm

By incorporating the nonlocal similarity regularization term into the LANR model in (17), we have

$$\begin{aligned} \hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{DHN}^H \mathbf{w}||_2^2 + \lambda_1 ||\mathbf{Gw}||_2^2 + \\ \lambda_2 \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{b}_i^T \boldsymbol{\beta}_i) \,\text{s.t.}\, 1^T \mathbf{w}_i = 1 \end{aligned} \tag{21}$$

where λ_2 is a constant balancing the contribution of nonlocal regularization. For the convenience of expression, we can rewrite $\sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{b}_i^T \boldsymbol{\beta}_i)$ as $||(\mathbf{I} - \mathbf{B})\mathbf{N}^H \mathbf{w}||_2^2$, where $\mathbf{I}$ is the identity matrix and $\mathbf{B}$ is defined as

$$\mathbf{B}(i,j) = \begin{cases} b_i^l, & \text{if } \mathbf{x}_i^l \text{ is an element of } \boldsymbol{\beta}_i, b_i^l \in \mathbf{b}_i; \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

Then, (21) can be rewritten as

$$\begin{aligned} \hat{\mathbf{w}} = \arg\min_{\mathbf{w}} ||\mathbf{y} - \mathbf{DHN}^H \mathbf{w}||_2^2 + \lambda_1 ||\mathbf{Gw}||_2^2 \\ + \lambda_2 ||(\mathbf{I} - \mathbf{B})\mathbf{N}^H \mathbf{w}||_2^2 \,\text{s.t.}\, 1^T \mathbf{w}_i = 1. \end{aligned} \tag{23}$$

In (23), the first term is the data fidelity term to ensure the solution $\hat{\mathbf{x}} = \mathbf{N}^H \hat{\mathbf{w}}$ can well fit the observation $\mathbf{y}$ after image degradation process by operators $\mathbf{B}$ and $\mathbf{D}$; the second term is the local geometry prior based adaptive regularization term to preserve the local geometry of the data manifold; the third term is the nonlocal similarity regularization term, which uses the nonlocal redundancy to enhance each local patch.

E. Optimization of LANR-NLM

In our objective function, there are two regularization terms, the nonlocal self-similarity regularization and the local geometry regularization. The former takes advantage of the redundancy of similar patches in natural images, while the latter can be used to regularize the modeling of the nonlinear relationship between LR and HR spaces.

Based on the above two considerations, we firstly apply the local geometry prior to regularize the patch representation to predict the HR target patch on a pixel-by-pixel basis. And then, we utilize the NLM filter to improve the super-resolved outcome. Therefore, the optimization of LANR-NLM includes the following two main steps: learning the patch representation (i.e., projection matrix) for each anchor patch in the dictionary, and refining the outcome by the NLM filter.

The solution of a regularized least square in (14) can be derived analytically as

$$\mathbf{w}_i = (\mathbf{N}_{i,j}^{L\,T} \mathbf{N}_{i,j}^L + \lambda_1 \mathbf{U})^{-1} \mathbf{N}_{i,j}^{L\,T} \mathbf{y}_i \tag{24}$$

where $\mathbf{U}$ is a $K \times K$ diagonal matrix with

$$\mathbf{U}_{kk} = g_{i,k}, k = 1, 2, ..., K. \tag{25}$$

In analogy to (11) and (12), we can obtain the stored projection matrices of our proposed LANR-NLM method

$$\mathbf{P}^j = \mathbf{N}_j^H (\mathbf{N}_{i,j}^{L\,T} \mathbf{N}_{i,j}^L + \lambda_1 \mathbf{U})^{-1} \mathbf{N}_{i,j}^{L\,T}. \tag{26}$$

If we calculate the projection matrix for each dictionary atom off-line (as given in Algorithm 1), we then can predict the HR image patch by patch. Integrating all the super-resolved HR patches according to their positions, we can generate the target HR image $\mathbf{x}$.

Given the estimated HR image, we adopt the gradient descent rule to obtain the optimal result as in [35]. The entire image SR process of the proposed method is given in Algorithm 2. In Algorithm 2, the LR and HR dictionaries are trained by the first six steps in Algorithm 1. e is a predefined scalar controlling the convergence of the iterative process, and $maxIter$ is the allowed maximum number of iterations. The step δ is set to 5. The iteration parameters are experimentally set as $M_0 = 20$ and $maxIter = 160$, respectively. Note that to reduce computational complexity, we update the NLM in every M_0 iterations.

IV. Experiments

In this section, we analyze the performance of our proposed LANR-NLM method in relation to its design parameters and compare it quantitatively and qualitatively with ANR and other state-of-the-art methods.

The peak signal to noise ratio (PSNR), structure similarity (SSIM) index [59] and feature similarity (FSIM) [60] are used to assess the objective quality of the reconstructed images. PSNR measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. SSIM is a method for measuring the similarity between two images. Compared with the measurement of PSNR, SSIM can better reflect the structure similarity between the target image and the reference image. SSIM is designed to improve the traditional evaluation methods such as PSNR and mean squared error (MSE), which have proven to be inconsistent with human eye perception. In addition to the well-known PSNR and SSIM numeric metrics, we also utilize one recently proposed image quality assessment approach, i.e., FSIM model.

It is based on the observation that human visual system perceives an image mainly according to its low-level features and is among the leading image quality assessment models in terms of prediction accuracy.

Since human visual system is more sensitive to the luminance component, the SR reconstruction is only performed on the luminance channel of color images, and the simple Bicubic interpolation is used for the chromatic components. Therefore, in all our experiments, the numeric metrics are measured only on the luminance component.

A. Experimental Configurations

1) Databases Training Set. For learning-basedSR methods, training set has a direct impact on the reconstruction quality of the super-resolved images. In our experiments, follow the same procedure in [27], [51], [58], exactly the same 91 example training images are used to generate example HR and LR patches.

Testing set: Following ANR [51], we use "Set5" and "Set14" which respectively contain 5 and 14 commonly used images for SR evaluation. Note that all the test images are absent completely in the training set.

2) Experimental Settings: In our experiments, the LR images are obtained from the original HR images by a 7×7 Gaussian blurring operator with a standard deviation of 1.6 and decimated by a factor of 3. We use the size of 9×9 pixels for HR image patches and the overlap between neighbor patches is 6 pixels. The size of the corresponding LR image patches is set to 3×3 pixels with an overlap of 2 pixels. Note that the overlap patch representation and reconstruction strategy is very time consuming, which can be accelerated by some parallel computation algorithms [61], [62].

We use the same feature extraction strategy in Yang *et al.* [27], Zeyde *et al.* [58] and Timofte *et al.* [51]. Specifically, to extract the high frequency information of LR images, four 1-D filters ($f_1 = [-1, 0, 1]$, $f_2 =, f_1^T$, $f_3 = [1, 0, -2, 0, 1]$, $f_4 = f_3^T$) are used to extract the derivatives of each patch, and then PCA dimensionality reduction is applied to project the extracted features onto a low-dimensional subspace while preserving 99.9% of the total variance. This usually leads to features of about 30 dimensions for up-scaling factor 3 and 3×3 LR patch size. As for the HR patch features, we subtract the Bicubic interpolated LR image from the HR image to create the normalized HR patches.

We adopt the same dictionary training method in [58] and [51]. The two dictionaries for HR and LR image patches are trained from 135,581 patch pairs randomly sampled from natural images collected from the 91 example training images. Firstly, we obtain a sparse dictionary $\mathbf{D}_L$ by optimizing over LR patches. And then, we reconstruct its corresponding $\mathbf{D}_H$ with the same coefficients obtained on the LR dictionary. Note that we choose the training set size to be 1024 as suggested in [51].

For our proposed LANR-NLM model, there are three parameters need to be set, i.e., λ_1, K, and α. The following reported

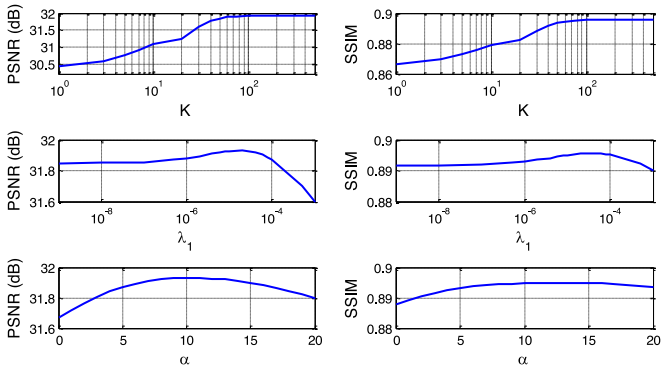


Fig. 2. Influence of the nearest neighbor number K, the regularization parameter λ_1, and the power parameter α of the proposed LANR-NLM method.

Fig. 3. SR outcome comparisons of LANR (left) and LANR-NLM (right).

SR results are obtained using the best parameters, $\lambda_1 = 1e-5$, $K = 200$, and $\alpha = 11$. λ_2 of the proposed LANR-NLM algorithm is set to 0.005. In the following, we will provide some analysis on the parameter settings.

B. Parameters Analysis

In this subsection, we analyze the influence of the nearest neighbor number K, the regularization parameter λ_1 and the power parameter α of the proposed LANR-NLM method. It is

TABLE I
SR PERFORMANCES OF DIFFERENT APPROACHES IN TERMS OF PSNR (DB), SSIM, AND FSIM PER IMAGE ON THE SET5 DATASET

Set5 Image	Bicubic			ScSR [27]			Zeyde [58]			ANR [51]			LANR			LANR-NLM		
	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
baby	29.43	0.8348	0.9309	29.36	0.8303	0.9246	34.43	0.9089	0.9863	34.71	0.9144	0.9885	34.97	0.9194	0.9906	35.10	0.9230	0.9918
bird	27.54	0.8369	0.8697	27.64	0.8391	0.8726	33.59	0.9341	0.9380	33.93	0.9399	0.9423	34.44	0.9432	0.9471	34.54	0.9458	0.9482
butterfly	20.72	0.7201	0.7443	21.22	0.7547	0.7748	25.24	0.8561	0.8274	25.37	0.8562	0.8257	25.92	0.8694	0.8322	26.01	0.8724	0.8360
head	30.21	0.7373	0.8400	30.04	0.7276	0.8236	33.11	0.8037	0.8955	33.30	0.8110	0.9021	33.56	0.8211	0.9139	33.65	0.8256	0.9178
woman	24.50	0.8060	0.8448	24.52	0.8073	0.8514	29.62	0.9024	0.9122	29.89	0.9069	0.9134	30.34	0.9118	0.9215	30.35	0.9123	0.9238
Average	26.48	0.7870	0.8459	26.56	0.7918	0.8494	31.20	0.8810	0.9119	31.44	0.8857	0.9144	31.85	0.8930	0.9211	31.93	0.8958	0.9235
Gain	5.45	0.1088	0.0776	5.37	0.1040	0.0741	0.73	0.0148	0.0116	0.49	0.0101	0.0091	0.14	0.0038	0.0028	–	–	–

TABLE II
SR PERFORMANCES OF DIFFERENT APPROACHES IN TERMS OF PSNR (DB), SSIM, AND FSIM PER IMAGE ON THE SET14 DATASET

Set14 Image	Bicubic			ScSR [27]			Zeyde [58]			ANR [51]			LANR			LANR-NLM		
	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
baboon	22.13	0.4339	0.8503	22.22	0.4485	0.8571	23.24	0.5518	0.9386	23.48	0.5871	0.9466	23.49	0.5901	0.9537	23.56	0.6038	0.9566
barbara	24.29	0.6608	0.8833	24.18	0.6597	0.8819	26.43	0.7615	0.9594	26.66	0.7748	0.9624	26.65	0.7756	0.9669	26.79	0.7842	0.9695
bridge	22.41	0.5130	0.8598	22.53	0.5314	0.8679	24.60	0.6588	0.9489	24.86	0.6892	0.9539	24.94	0.6920	0.9607	25.03	0.7031	0.9635
coastguard	25.01	0.5162	0.6690	25.12	0.5176	0.6766	26.73	0.6191	0.7660	26.98	0.6428	0.7783	26.99	0.6433	0.7992	27.00	0.6437	0.8063
comic	20.83	0.5537	0.7133	21.02	0.5839	0.7345	23.42	0.7174	0.8069	23.78	0.7432	0.8142	23.97	0.7531	0.8276	24.09	0.7632	0.8324
face	30.17	0.7358	0.8402	30.01	0.7268	0.8240	33.12	0.8038	0.8966	33.47	0.8181	0.9034	33.58	0.8207	0.9143	33.66	0.8250	0.9181
flowers	24.25	0.6955	0.7778	24.49	0.7103	0.7899	27.73	0.8120	0.8561	28.11	0.8299	0.8627	28.42	0.8360	0.8743	28.56	0.8417	0.8783
foreman	27.29	0.8551	0.8692	27.28	0.8570	0.8753	32.29	0.9165	0.9240	32.49	0.9194	0.9284	33.41	0.9292	0.9408	33.43	0.9298	0.9315
lenna	27.82	0.7887	0.9181	27.82	0.7861	0.9140	32.28	0.8639	0.9802	32.67	0.8742	0.9837	32.97	0.8768	0.9872	33.14	0.8807	0.9887
man	24.56	0.6500	0.8818	24.66	0.6548	0.8821	27.45	0.7626	0.9627	27.66	0.7784	0.9670	27.93	0.7852	0.9736	28.01	0.7913	0.9755
monarch	25.90	0.8635	0.9135	26.25	0.8695	0.9199	30.37	0.9283	0.9744	30.62	0.9313	0.9770	31.06	0.9359	0.9808	31.23	0.9378	0.9833
pepper	28.20	0.8193	0.9308	28.14	0.8119	0.9255	33.45	0.8773	0.9851	33.25	0.8782	0.9852	33.83	0.8821	0.9881	33.96	0.8848	0.9893
ppt3	20.88	0.8020	0.8297	21.08	0.8075	0.8376	24.61	0.8960	0.9300	24.58	0.8832	0.9258	25.12	0.9024	0.9393	25.15	0.9011	0.9432
zebra	22.35	0.6445	0.8299	22.51	0.6585	0.8393	27.51	0.8091	0.9568	27.85	0.8299	0.9588	28.10	0.8338	0.9656	28.28	0.8425	0.9697
Average	24.72	0.6809	0.8405	24.81	0.6874	0.8447	28.09	0.7841	0.9204	28.32	0.7985	0.9248	28.61	0.8040	0.9337	28.71	0.8095	0.9361
Gain	3.99	0.1286	0.0956	3.90	0.1221	0.0914	0.62	0.0254	0.0157	0.39	0.0110	0.0113	0.10	0.0055	0.0024	–	–	–

worth mentioning that we only conduct experiments on Set5, and we can obtain similar conclusions on Set14.

1) Influence of the Nearest Neighbor Number K: The first row of Fig. 2 shows the plots of the average PSNR (dB) and SSIM values of all 5 test images from Set5 according to different values of nearest neighbor number. We can see that this parameter has a large impact on the performance of the proposed method: as the nearest neighbor number increases, the gain of the proposed method becomes larger, which implies that large training set will induce more detailed features and good reconstruction results. Therefore, we set K to 200 in our experiments.

2) Influence of the Regularization Parameter λ: In the second row of Fig. 2, we present the average PSNR (dB) and SSIM values of all 5 test images in Set5 according to various values for the regularization parameter λ, which controls the weight of locality in the objective function. We can find that when $\lambda = 0$, the performance of LANR-NLM is not the best. With the increase of locality, the SR performance is improved. This implies that the locality constraint is essential for regression reconstruction. However, we also observe that the value of λ could not be set too large because the reconstruction error in the objective function cannot be ignored. Therefore, LANR-NLM can achieve good results by selecting a proper regularization parameter λ (e.g., $\lambda = 1e-5$),

3) Influence of the Power Parameter α: In the third row of Fig. 2, we show the average PSNR (dB) and SSIM values of all 5 test images in Set5 according to various values for the regularization parameter α, which controls the weight decay speed of the locality adaptor. When α is set to around 11, LANR-NLM achieves the best performance. This implies that by setting a relative large value of α, samples that are far away from the observation sample will be heavily penalized and samples that are close to the observation sample will be given more freedom. This implies LANR-NLM can well explore the nonlinear manifold structure of the data space.

C. Results Comparison

To demonstrate the effectiveness of the newly introduced NLM prior, we report some visual results of the *bird* and *lenna* of LANR and LANE-NLM in Fig. 3. Based on the results in Fig. 3, we find that NLM prior does affect the SR reconstruction results. This is principally because local constraint and nonlocal similarity in natural images can complement each other. The

Fig. 4. Visual reconstruction results for part of the image "woman" from Set5. Top (from left to right): ground truth HR image, reconstructed HR image by Bicubic interpolation, and Yang *et al.* [27]. Bottom (from left to right): reconstructed HR image by Zeyde *et al.* [58], Timofte *et al.* [51], and our proposed LANR-NLM.

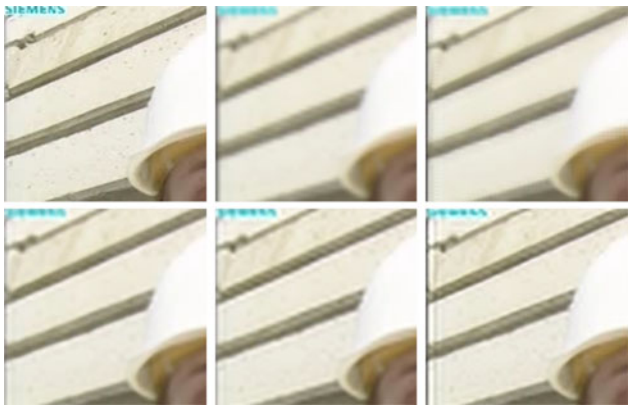


Fig. 5. Visual reconstruction results for part of the image "foreman" from Set14. Top (from left to right): ground truth HR image, reconstructed HR image by Bicubic interpolation, and Yang *et al.* [27]. Bottom (from left to right): reconstructed HR image by Zeyde *et al.* [58], Timofte *et al.* [51], and our proposed LANR-NLM.

Fig. 6. Visual reconstruction results for part of the image "monarch" from Set14. Top (from left to right): ground truth HR image, reconstructed HR image by Bicubic interpolation, and Yang *et al.* [27]. Bottom (from left to right): reconstructed HR image by Zeyde *et al.* [58], Timofte *et al.* [51], and our proposed LANR-NLM.

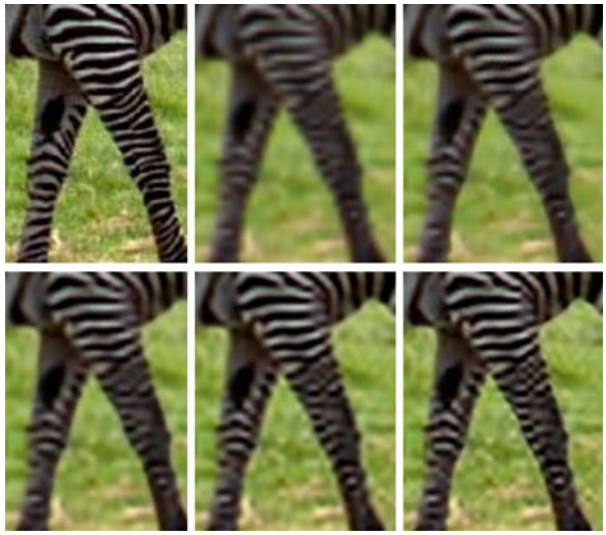


Fig. 7. Visual reconstruction results for part of the image "zebra" from Set14. Top (from left to right): ground truth HR image, reconstructed HR image by Bicubic interpolation, and Yang *et al.* [27]. Bottom (from left to right): reconstructed HR image by Zeyde *et al.* [58], Timofte *et al.* [51], and our proposed LANR-NLM.

combination of these two priors can contribute more faithful and reliable SR reconstruction.

In addition, we also compare our proposed LANR-NLM method to the sparse coding based algorithms (Yang *et al.* [27] and Zeyde *et al.* [58] and ANR [51]. The results of the Bicubic interpolation method are also given as baselines for comparison. The results of the comparison algorithms are obtained using the corresponding software packages that are publicly available.[1]

Tables I and II compare the image reconstruction performances (PSNR (dB), SSIM and FSIM) of various methods under the same testing conditions. We can see that the proposed LANR-NLM method reaches the highest numeric results in all experiments. The gain is 0.49 dB in term of PSNR, 0.0101 in term of SSIM, and 0.0091 in term of FSIM better than ANR [51] on Set5, and 0.39 dB in term of PSNR, 0.0110 in term of SSIM, and 0.0113 in term of FSIM better than ANR [51] on Set14, respectively. When compared to other approaches, the gains of our proposed method are much more obvious on both test sets (Set5 and Set14). To demonstrate the effectiveness of the NLM constraint, we also report the results with and without NLM, i.e., LANR-NLM and LANR. Results on both test sets all show that LANR-NLM is slightly better than LANR.

In Figs. 4–7, we can see that our proposed method achieves slightly better quality over the other methods on a couple of images. The reconstructed images of our method are much more visually pleasing. Specifically, our prosed method can gain sharper edges. For example, the grid lines of the head scarf in the *woman* image (Fig. 4), the diagonal lines in the background of the *foreman* image (Fig. 5), the wing of

[1][Online]. Available: http://www.ifp.illinois.edu/jyang29/codes/ScSR.rar; http://www.cs.technion.ac.il/~elad/Various/Single_Image_SR.zip; http://www.vision.ee.ethz.ch/~timofter/software/SR_NE_ANR.zip

the butterfly in the *monarch* image (Fig. 6), and the foreleg in the *zebra* image (Fig. 7).

V. Conclusion

This paper presents a novel image SR method called locally regularized anchored neighborhood regression with nonlocal means (LANR-NLM). It applies locality constraint to select similar dictionary atoms and assigns different freedom to each dictionary atom according to its correlation to the input LR patch. By introducing this flexible prior, the proposed method can well learn simple regression functions and generate natural looking results with sharp edges and rich textures. Furthermore, we also incorporate the nonlocal self-similarity to the our proposed model to refine the outcome. Experimental results on standard benchmarks with qualitative and quantitative comparisons against several state-of-the-art SR methods demonstrate the effectiveness of the proposed LANR-NLM approach. Particularly, the improvement of LANR-NLM over the ANR method [51] (which is considered as the current state-of-the-art SR algorithm) is around 0.4 dB in terms of PSNR, 0.01 in terms of SSIM and 0.01 in terms of FSIM.

Acknowledgment

The authors would like to thank the anonymous reviewers for their invaluable comments and constructive suggestions.

References

[1] L. Wang, K. Lu, and P. Liu, "Compressed sensing of a remote sensing image based on the priors of the reference image," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 736–740, Apr. 2015.

[2] H. Lu *et al.*, "Reference information based remote sensing image reconstruction with generalized nonconvex low-rank approximation," *Remote Sens.*, vol. 8, no. 6, p. 499, 2016.

[3] H. Greenspan, "Super-resolution in medical imaging," *Comput. J.*, vol. 52, no. 1, pp. 43–63, 2009.

[4] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug. 2014.

[5] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4220–4231, Oct. 2014.

[6] J. Jiang, C. Chen, K. Huang, Z. Cai, and R. Hu, "Noise robust position-patch based face super-resolution via Tikhonovregularized neighbor representation," *Inf. Sci.*, vols. 367/368, pp. 354–372, 2016.

[7] J. Jiang *et al.*, "SRSLP: A face image super-resolution algorithm using smooth regression with local structure prior," *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2016.2601020.

[8] R. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, no. 2, pp. 317–339, 1984.

[9] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single image super-resolution via coupled constraint," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 469–480, Feb. 2012.

[10] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[11] X. Zhang and X. Wu, "Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, Jun. 2008.

[12] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[13] X. Li, H. He, R. Wang, and D. Tao, "Single image superresolution via directional group sparsity and directional features," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2874–2888, Sep. 2015.

[14] Y. Zhu, K. Li, and J. Jiang, "Video super-resolution based on automatic key-frame selection and feature-guided variational optical flow," *Signal Process., Image Commun.*, vol. 29, no. 8, pp. 875–886, 2014.

[15] K. Li, Y. Zhu, J. Yang, and J. Jiang, "Video super-resolution using an adaptive superpixel-guided auto-regressive model," *Pattern Recog.*, vol. 51, pp. 59–71, 2016.

[16] 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.

[17] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.

[18] M. Elad and A. Feuer, "Superresolution restoration of an image sequence: Adaptive filtering approach," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 387–395, Mar. 1999.

[19] J. Ma *et al.*, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.

[20] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.

[21] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, 2016.

[22] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[23] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 83–97, Jan. 2004.

[24] 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.

[25] J. Jiang, X. Ma, Z. Cai, and R. Hu, "Sparse support regression for image super-resolution," *IEEE Photon. J.*, vol. 7, no. 5, pp. 1–11, Oct. 2015.

[26] Y. Hu, N. Wang, D. Tao, X. Gao, and X. Li, "Serf: A simple, effective, robust, and fast image super-resolver from cascaded linear regression," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4091–4102, Sep. 2016.

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

[28] Z. Wang, R. Hu, S. Wang, and J. Jiang, "Face hallucination via weighted adaptive sparse regularization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 802–813, May. 2014.

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

[30] Z. Xiong, D. Xu, X. Sun, and F. Wu, "Example-based super-resolution with soft information and decision," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1458–1465, Oct. 2013.

[31] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, May. 2012.

[32] S. Hawe, M. Kleinsteuber, and K. Diepold, "Analysis operator learning and its application to image reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2138–2150, Jun. 2013.

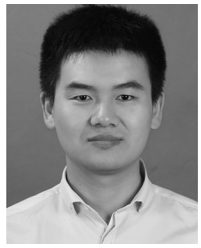
[33] J. Ren, J. Liu, and Z. Guo, "Context-aware sparse decomposition for image denoising and super-resolution," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1456–1469, Apr. 2013.

[34] Y. Zhang, J. Liu, W. Yang, and Z. Guo, "Image super-resolution based on structure-modulated sparse representation," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2797–2810, Sep. 2015.

[35] W. Dong, D. 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.

[36] W. Dong, X. Wu, and G. Shi, "Sparsity fine tuning in wavelet domain with application to compressive image reconstruction," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5249–5262, Dec. 2014.

[37] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2223–2231.
[38] J. Wang *et al.*, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 3360–3367.
[39] C. Chen and J. E. Fowler, "Single-image super-resolution using multihypothesis prediction," in *Proc. 46th Asilo mar Conf. Signals, Syst., Comput.*, Nov. 2012, pp. 608–612.
[40] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Mach. Learn. Res.*, vol. 4, pp. 119–155, 2003.
[41] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1674–1684, Sep. 2016.
[42] 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, 2000.
[43] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, Jun. 2005, pp. 60–65.
[44] P. Arias, V. Caselles, and G. Sapiro, "A variational framework for non-local image inpainting," in *Proc. 7th Int. Conf. Energy Minimization Methods Comput. Vis. Pattern Recog.*, 2009, pp. 345–358.
[45] L. Liu, L. Chen, C. L. P. Chen, Y. Y. Tang, and C. M. Pun, "Weighted joint sparse representation for removing mixed noise in image," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2016.2521428.
[46] C. L. P. Chen, L. Liu, L. Chen, Y. Y. Tang, and Y. Zhou, "Weighted couple sparse representation with classified regularization for impulse noise removal," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4014–4026, Nov. 2015.
[47] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep.-Oct. 2009, pp. 2272–2279.
[48] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1382–1394, Apr. 2013.
[49] M.-C. Yang and Y.-C. Wang, "A self-learning approach to single image super-resolution," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 498–508, Apr. 2013.
[50] Z. Zhu, F. Guo, H. Yu, and C. Chen, "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec. 2014.
[51] R. Timofte, V. De, 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.
[52] J. Jiang, J. Fu, T. Lu, R. Hu, and Z. Wang, "Locally regularized anchored neighborhood regression for fast super-resolution," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2015, pp. 1–6.
[53] A. N. Tikhonov and V. Arsenin, *Solutions of ill-posed problems*. Toronto, ON, Canada: Vh Winston, 1977.
[54] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, Nonlinear Phenom.*, vol. 60, no. 1, pp. 259–268, 1992.
[55] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[56] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Vis. Pattern Recog.*, Jun.-Jul. 2004, vol. 1, pp. 275–282.
[57] M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. 23rd Brit. Mach. Vis. Conf.*, 2012, pp. 1–10.
[58] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations." Berlin, Germany: Springer, 2012, vol. 6920, pp. 711–730.
[59] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[60] 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.
[61] D. Chen *et al.*, "Parallel simulation of complex evacuation scenarios with adaptive agent models," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 3, pp. 847–857, Mar. 2015.
[62] D. Chen *et al.*, "Fast and scalable multi-way analysis of massive neural data," *IEEE Trans. Comput.*, vol. 64, no. 3, pp. 707–719, Mar. 2015.



Junjun Jiang (M'15) received the B.S. degree from the School of Mathematical Sciences, Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree from the School of Computer, Wuhan University, Wuhan, China, in 2014.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. He has been a Project Researcher with the National Institute of Informatics, Tokyo, Japan, since 2016. He has authored and coauthored more than 60 scientific articles and holds eight Chinese patents. His research interests include image processing and computer vision.

Xiang Ma received the B.S. degree from the North China Electric Power University, Beijing, China, in 1999, and the Ph.D. degree from Xian Jiaotong University, Xi'an, China, in 2011.

He is currently an Associate Professor with the School of Information Engineering, Changan University, Xi'an, China. His research interests include face image processing and image processing in machine and intelligent transportation.

Chen Chen received the B.E. degree in automation from the Beijing Forestry University, Beijing, China, in 2009, the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2012, and the Ph.D. degree from the University of Texas at Dallas, Richardson, TX, USA, in 2016.

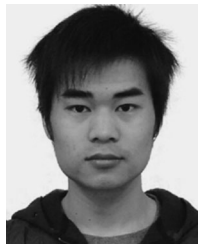
He is currently a Post Doctoral with the Center for Research in Computer Vision, University of Central Florida, Orlando, FL, USA. He has authored or coauthored more than 40 papers in refereed journals and conferences. His research interests include compressed sensing, signal and image processing, pattern recognition, and computer vision.

Tao Lu received the B.S. and M.S. degrees from the Wuhan Institute of Technology, Wuhan, China, in 2003 and 2008, respectively, and the Ph.D. degree from the National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan, China, in 2013.

He is currently an Associate Professor with the School of Computer Science and Engineering,Wuhan Institute of Technology, a Research Member with the Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan, China, and a Visiting Scholar with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His research interests include image/video processing, computer vision, and artificial intelligence.

Zhongyuan Wang (M'13) received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008.

He is currently a Professor with the School of Computer, Wuhan University, Wuhan, China. He is currently directing two projects funded by the National Natural Science Foundation Program of China. His research interests include video compression, image processing, and multimedia communications.



Jiayi Ma (M'16) received the B.S. and Ph.D. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He is currently an Associate Professor with the Electronic Information School, Wuhan University, Wuhan, China, where he was a Post Doctoral from 2014 to 2015. His current research interests include computer vision, machine learning, and pattern recognition.