Coupled Deep Autoencoder for Single Image Super-Resolution

Kun Zeng, Jun Yu, *Member, IEEE*, Ruxin Wang, Cuihua Li, and Dacheng Tao, *Fellow, IEEE*

***Abstract*—Sparse coding has been widely applied to learning-based single image super-resolution (SR) and has obtained promising performance by jointly learning effective representations for low-resolution (LR) and high-resolution (HR) image patch pairs. However, the resulting HR images often suffer from ringing, jaggy, and blurring artifacts due to the strong yet *ad hoc* assumptions that the LR image patch representation is equal to, is linear with, lies on a manifold similar to, or has the same support set as the corresponding HR image patch representation. Motivated by the success of deep learning, we develop a data-driven model coupled deep autoencoder (CDA) for single image SR. CDA is based on a new deep architecture and has high representational capability. CDA simultaneously learns the intrinsic representations of LR and HR image patches and a big-data-driven function that precisely maps these LR representations to their corresponding HR representations. Extensive experimentation demonstrates the superior effectiveness and efficiency of CDA for single image SR compared to other state-of-the-art methods on Set5 and Set14 datasets.**

***Index Terms*—Autoencoder, deep learning, single image super-resolution (SR).**

I. Introduction

SINGLE image super-resolution (SR) [1]–[5] aims to generate a visually pleasing high-resolution (HR) image from a single low-resolution (LR) input. This important yet difficult topic has gained more attention over recent years as the demand for high-quality images in vision systems has increased. The key problems in single image SR include: 1) prior knowledge on the lost high frequency details is difficult to explore and 2) how LR and HR images correspond to each other is often ambiguous or even arbitrary. A number of SR methods have been developed that can be categorized into three main groups.

1) Interpolation-based SR. These methods produce an HR image from a given LR image by estimating the pixels in the HR grids via conventional interpolation algorithms such as bi-linear or bi-cubic interpolation. These methods are simple but prone to producing blurred edges and unpleasant artifacts.
2) Reconstruction-based SR. These methods incorporate specific prior knowledge on the HR details. Several priors have been used to restore HR images, e.g., the edge prior [6], the gradient profile prior [7], and non-local means [8]. These approaches preserve edges but do not reconstruct plausible details effectively at high magnification.
3) Learning-based SR. These methods assume that the high-frequency details lost at the decreased resolution can be predicted by learning an SR model from an external LR and HR image pair dataset. Learning-based image processing [9], including learning-based SR, always explores image pair priori which describes the dependency between training image pairs by image pair analysis [10], [11]. These methods have shown impressive performance for single image SR [12]–[28].

Recently, sparse coding (SC) [13], [29]–[31], an important learning-based method, has been studied for single image SR [16]–[18], [32]–[37] with promising results. A conventional assumption made when designing SC schemes is that the sparse representation of the LR image patch over the LR dictionary is identical to that of the corresponding HR image patch over the HR dictionary [13]. It may fail to produce pleasant SR results, because this assumption is too restrictive or ineffective in general.

To overcome this assumption, Dong *et al.* [38] introduced SC noise, which allows the sparse representation of the LR patch to be a slightly corrupted representation of the corresponding HR patch. By employing nonlocal means, the HR representation can be well estimated based on a set of LR representations. Similarly, Wang *et al.* [32] imposed a linear relationship to the sparse representations of the LR/HR image patch pair such that the representation of the HR image patch can be linearly transformed from that of the LR patch. Yang and Yang [15] also employed a similar method and learned a set of simple linear functions for each patch cluster such that the HR patch can be reconstructed via these functions.

Manuscript received March 29, 2015; revised October 26, 2015; accepted November 11, 2015. Date of publication November 26, 2015; date of current version December 14, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61472110 and Grant 61373077, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR15F020002, in part by the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant 20110121110020, in part by the National Defense Basic Scientific Research Program of China under Grant B-0110155, and in part by the National Defense Science and Technology Key Laboratory Foundation under Grant 9140C-30211ZS-8. This paper was recommended by Associate Editor X. You. *(Corresponding author: Jun Yu.)*

K. Zeng and C. Li are with Computer Science Department, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China (e-mail: zengkun301@aliyun.com; chli@xmu.edu.cn).

J. Yu is with the School of Computer Science and the Key Laboratory of Complex Systems Modeling and Simulation, Ministry of Education, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: yujun@hdu.edu.cn).

R. Wang and D. Tao are with the Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: rosinwang@gmail.com; dacheng.tao@uts.edu.au).

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/TCYB.2015.2501373

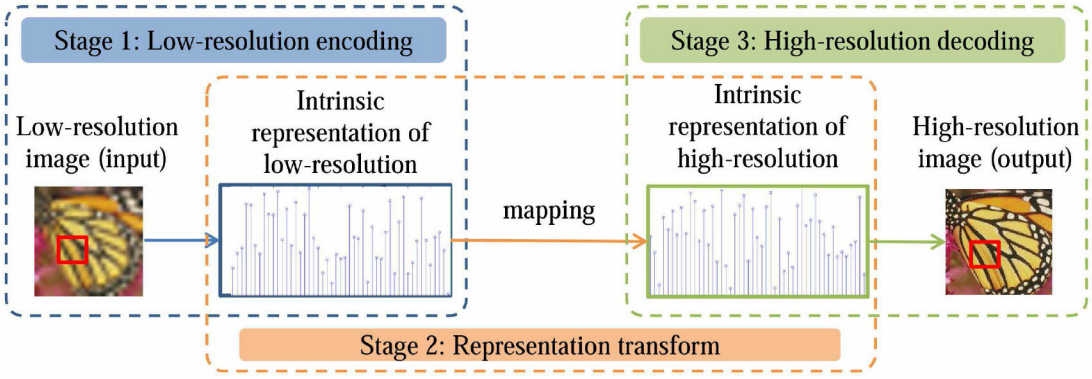


Fig. 1. Flowchart showing how CDA achieves single image SR. The process has three stages. Left: stage 1 obtains the intrinsic representations of the LR image patches. Middle: stage 2 transforms the LR intrinsic representations to the corresponding HR intrinsic representations. Right: stage 3 reconstructs the HR image patches based on the corresponding intrinsic representations.

Considering the nonlinearity of many image patches, the manifold assumption proposed by Chang *et al.* [12] highlights that the manifolds of LR image patches and their corresponding HR patches are locally consistent in their respective feature spaces. Inspired by locally linear embedding (LLE) [39], neighbor embedding (NE) [12] facilitates the prediction of the HR patch as a linear combination of its HR neighbors corresponding to the LR neighbors.

Instead of imposing identical representations on LR/HR image patches, Jia *et al.* [40] proposed a coupled SC, which requires that the sparse representations of the LR/HR image patches share the same support set.

The aforementioned methods typically rely on the same four-step procedure: 1) extracting the first- and second-order features from the overlapping LR patches; 2) encoding the features based on an LR dictionary; 3) mapping the LR representations to the HR representations; and 4) reconstructing HR patches based on an HR dictionary. During this process, however, seeking the sparse representation of LR patches requires an L_1 problem to be solved, which is inefficient. Furthermore, the mapping between the LR/HR sparse representations is critical to the outcome. The assumptions used—identical mapping [13], linear mapping [15], [32], manifold similarity [12], and the same support set [40]—are strong and *ad hoc* and may fail when the conditions are inexact, resulting in ringing, jaggy, and blurring artifacts in the reconstructed HR images. More effective and efficient feature encoding and mapping learning methods are urgently required for real-world applications.

Deep learning architectures have been useful for discovering hidden representations in natural images and have proven success in various vision tasks. In this big data era, the extensive availability of training images enable deep models to be generic and flexible and avoid the need to manually design task-specific image features [41], [42]. Therefore, learning is the key in this active research area. Inspired by their learning capability and capacity to discover representations, we hypothesize that deep architectures would be ideally suited to seek the proper representations for HR/LR image pairs and modeling their relationship. For example, autoencoder [43], [44] is an unsupervised feature-learning scheme in which the internal layer acts as a generic extractor of intrinsic image representations. A multilayer perceptron provides a double-layer structure that can efficiently map the input data onto appropriate outputs. In addition, deep learning can exploit parallel graphics processing unit computation and deliver high speeds in the forward pass. These advantages make deep models an attractive option for handling the single image SR problem.

In this paper, we develop a big-data-driven strategy for SR. Specifically, we design a new deep architecture called coupled deep autoencoder (CDA) for single image SR. CDA employs two autoencoders to extract hidden representations in LR/HR image patches, which we term "intrinsic representations" [45]. Since autoencoder encoding and decoding only involve multiplication, addition, and activation operations, intrinsic representation inference and HR patch reconstruction will be very fast. In learning the mapping function, we incorporate a one-layer neural network to map the intrinsic representations of LR patches to those of HR patches (see Fig. 1). A notable characteristic of this approach is that the neural network can simulate arbitrary relationships between the intrinsic representations of LR and HR patches in both linear and nonlinear cases; it is therefore more flexible than existing SC-based methods. Moreover, CDA is supervised and task-specific, and it can jointly explore the intrinsic representations of LR/HR data and the appropriate relationship between them. We demonstrate the effectiveness and efficiency of CDA by extensive experimentation on two popular SR datasets.

The remainder of this paper is organized as follows. CDA is presented in Section II. The experimental results on single image SR and comparisons with other state-of-the-art methods are presented in Section III. We conclude in Section IV.

II. Coupled Deep Autoencoder for SR

Autoencoder has recently shown promise in unsupervised feature learning. It aims to learn a compact representation of the input while retaining the most important information. This representation is expected to completely reconstruct the original input [43]. This property makes autoencoder useful for image SR by enabling us to seek the representations from LR data and use them to reconstruct the HR data. Formally,

given a set of samples $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$, where $\mathbf{y}_i \in \mathbb{R}^d$, the training objective of an autoencoder is to minimize the reconstruction error

$$\ell = \sum_i \left\| \mathbf{y}_i - \hat{\mathbf{y}}_i \right\|^2 \tag{1}$$

where $\mathbf{y}_i$ and $\hat{\mathbf{y}}_i$ are the original input and the reconstructed input, respectively. The hidden layer implies an encoding process and a decoding process

$$\begin{cases} \mathbf{h}_i = f\left(\mathbf{W}\mathbf{y}_i + \mathbf{b}\right) \\ \hat{\mathbf{y}}_i = f\left(\mathbf{W}^{'}\mathbf{h}_i + \mathbf{b}^{'}\right) \end{cases} \tag{2}$$

where $\mathbf{h}_i \in \mathbb{R}^n$ is the compact representation, $\mathbf{W}$ and $\mathbf{W}^{'}$ represent the weight matrices for encoding and decoding layers, $\mathbf{b}$ and $\mathbf{b}^{'}$ denote the bias terms. $f(\cdot)$ is the activation function, which we set as the sigmoid function in this paper

$$f(z) = \frac{1}{1 + \exp(-z)}. \tag{3}$$

The autoencoders can induce very useful representations of the inputs. However, they can only handle a single sample and cannot model the relationship between a sample pair. In image SR, we are interested in the joint task of discovering suitable representations for image pairs and encoding their relationship. We argue that a better representation should depend on not only the input image but also the internal relationship between the HR/LR image pairs. With this in mind, we develop the CDA.

A. CDA

CDA has a three-stage architecture, as shown in Fig. 1. The first and third stages employ two autoencoders for learning the representations of LR and HR image patches, respectively. The second stage incorporates a one-layer neural network to transform the LR representation into the HR representation.

Following the above notations, the two autoencoders generate the hidden representations $\mathbf{h}^L$ and $\mathbf{h}^H$, which we term the intrinsic representations of the LR and HR input, respectively. Given the LR input $\mathbf{y}_i$ and the corresponding HR input $\mathbf{x}_i$, the intrinsic representations can be obtained by

$$\mathbf{h}_i^L = f\left(\mathbf{W}_1\mathbf{y}_i + \mathbf{b}_1\right) \tag{4}$$

$$\mathbf{h}_i^H = f\left(\mathbf{W}_3^{'}\mathbf{x}_i + \mathbf{b}_3^{'}\right). \tag{5}$$

For reconstruction, the decoding processes imply that

$$\hat{\mathbf{y}}_i = f\left(\mathbf{W}_1^{'}\mathbf{h}_i^L + \mathbf{b}_1^{'}\right) \tag{6}$$

$$\hat{\mathbf{x}}_i = f\left(\mathbf{W}_3\mathbf{h}_i^H + \mathbf{b}_3\right). \tag{7}$$

The parameters $(\mathbf{W}_1, \mathbf{W}_1^{'}, \mathbf{b}_1, \mathbf{b}_1^{'})$ characterize the LR autoencoder (LRAE) while $(\mathbf{W}_3^{'}, \mathbf{W}_3, \mathbf{b}_3^{'}, \mathbf{b}_3)$ parameterize the HR autoencoder.

After obtaining the LR/HR intrinsic representations, the neural network implements mapping from $\mathbf{h}^L$ to $\mathbf{h}^H$. Mathematically, let us denote the parameters in this stage as $(\mathbf{W}_2, \mathbf{b}_2)$, where $\mathbf{W}_2$ is the weight matrix and $\mathbf{b}_2$ is the bias term. The mapping function then becomes

$$\mathbf{h}_i^H = f\left(\mathbf{W}_2\mathbf{h}_i^L + \mathbf{b}_2\right). \tag{8}$$

Algorithm 1 CDA for SR

Input: a LR image $\mathbf{Y}$ and well trained CDA model $\Theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$
Output: the HR image $\hat{\mathbf{X}}$
Step 1: Extract low-resolution image patches $\mathbf{y}_i$ using (9);
Step 2: for each image patch $\mathbf{y}_i$
Step 2.1: Obtain the LR intrinsic representation $\mathbf{h}_i^L$ by (4);
Step 2.2: Obtain the HR intrinsic representation $\mathbf{h}_i^H$ by (8);
Step 2.3: Obtain the HR image patch $\hat{\mathbf{x}}_i$ by (7);
Step 3: Reconstruct the HR image $\hat{\mathbf{X}}$ using (10).

The construction of the CDA suggests that the model is simple and flexible. The autoencoders ensure that the intrinsic representations are well fit to the LR and HR images and the neural network can learn complex relationships between the LR/HR representations; notably, the mapping function and the intrinsic representations are jointly optimized and thus correlated. Therefore, the constructed architecture is a data-driven model for single image SR. Note that we can replace autoencoder with the stacked autoencoder [46] or the de-noising autoencoder [47] to obtain further performance improvement.

B. Super-Resolution by CDA

For single image SR, CDA is a three-layer forward network employing a fast feed-forward process, as shown in Fig. 1. The SR steps are as follows.

Following the preprocessing step found in most SR methods, a single LR image is first upscaled to the desired size using bi-cubic interpolation. To avoid confusion, this interpolated LR image is denoted by $\mathbf{Y}$. The LR image patches $\mathbf{y}_i (i = 1, 2, \ldots, N)$ are obtained through

$$\mathbf{y}_i = \mathbf{R}_i\mathbf{Y} \tag{9}$$

where $\mathbf{R}_i$ is the operator to extract the ith local patch in $\mathbf{Y}$. Taking $\mathbf{y}_i$ as the input of CDA, the forward process incorporates (4), (7) and (8) to infer the LR intrinsic representation $\mathbf{h}_i^L$, the HR intrinsic representation $\mathbf{h}_i^H$, and the final restored HR patch $\hat{\mathbf{x}}_i$, respectively. To estimate the whole HR image $\hat{\mathbf{X}}$, we merge all restored patches by averaging the overlapping regions between adjacent patches

$$\hat{\mathbf{X}} = \left(\sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1} \sum_i \mathbf{R}_i^T\hat{\mathbf{x}}_i. \tag{10}$$

Algorithm 1 describes CDA for SR in detail. The dimensions of the hidden units in each layer are discussed in Section III.

C. Training CDA

CDA needs to discover the LR/HR intrinsic representations and simultaneously joint them using a well-trained mapping function. For this purpose, we have designed a two-part training procedure: the first part is initialization (stages 1–3 in Fig. 2), and the second part is the fine-tuning implemented in stage 4.

1) Initialization: To train CDA, the intrinsic representations of the LR/HR inputs are first generated. According to the autoencoder introduced in the beginning of Section II,

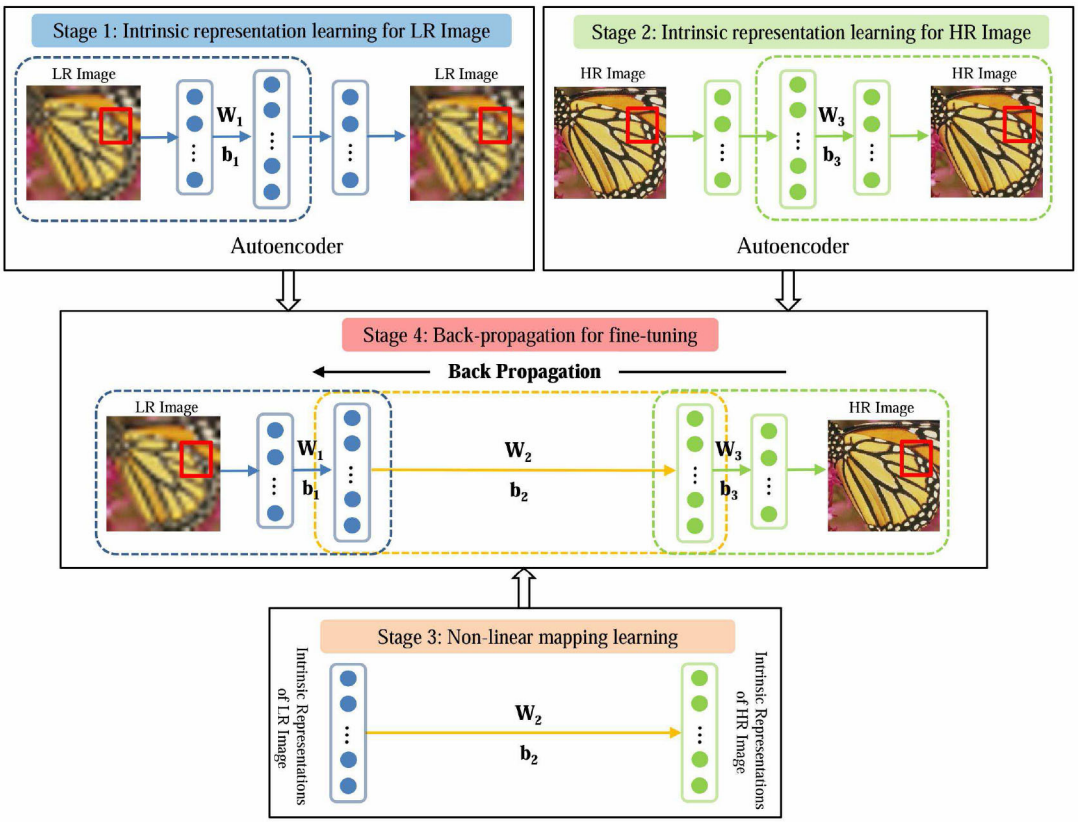


Fig. 2. Flowchart of the CDA training. In stage 1, we adopt a three-layer autoencoder to learn the intrinsic representations of the LR image patches; these are used to initialize the encoding layer of CDA. In stage 2, we seek the intrinsic representations of the HR image patches, which are used to initialize the decoding layer of CDA. In stage 3, we initialize the mapping layer of CDA by optimizing the one-layer neural network. In stage 4, the parameters of CDA are fine-tuned by back-propagation.

recall that, given an input LR image patch $\mathbf{y}_i$ and based on (4) and (6), the LR intrinsic representation $\mathbf{h}_i^L$ can be obtained by minimizing

$$\ell^L = \sum_i \left\|\mathbf{y}_i - \hat{\mathbf{y}}_i\right\|^2 \tag{11}$$

which is the reconstruction error of the LRAE. Similarly, for the HR counterpart (HRAE), given an HR image patch $\mathbf{x}_i$, the corresponding intrinsic representation $\mathbf{h}_i^H$ is induced by minimizing

$$\ell^H = \sum_i \left\|\mathbf{x}_i - \hat{\mathbf{x}}_i\right\|^2. \tag{12}$$

Based on the obtained $\mathbf{h}_i^L$ and $\mathbf{h}_i^H$, the optimization of the neural network in stage 3 learns the mapping function by minimizing

$$\sum_i \left\|\mathbf{h}_i^H - f\left(\mathbf{W}_2\mathbf{h}_i^L + \mathbf{b}_2\right)\right\|^2. \tag{13}$$

Combining (11)–(13), the whole objective function in initialization is obtained

$$\ell = \sum_i \left\|\mathbf{y}_i - \hat{\mathbf{y}}_i\right\|^2 + \sum_i \left\|\mathbf{x}_i - \hat{\mathbf{x}}_i\right\|^2 + \sum_i \left\|\mathbf{h}_i^H - f\left(\mathbf{W}_2\mathbf{h}_i^L + \mathbf{b}_2\right)\right\|^2. \tag{14}$$

To minimize (14), we first optimize the two autoencoders, after which the mapping function is learned based on the obtained intrinsic representations. Note that the LRAE and HRAE do not exchange data, thus optimizing $(\mathbf{W}_1, \mathbf{b}_1)$ and $(\mathbf{W}_3, \mathbf{b}_3)$ can be performed in parallel. When training each set of parameters, gradient-based methods can be used and perform well due to the convexity.

It is worth mentioning that until now, the parameters $(\mathbf{W}_1, \mathbf{b}_1)$ and $(\mathbf{W}_3, \mathbf{b}_3)$ are only used to encode and decode the respective input patches, respectively, while $(\mathbf{W}_2, \mathbf{b}_2)$ only encodes the mapping from $\mathbf{h}_i^L$ to $\mathbf{h}_i^H$. To realize CDA, next we discuss how to couple these parameters.

2) Fine-Tuning: The above initialization process does not guarantee that the obtained parameters are optimal for single image SR, since individually training each set of parameters may lead to a high reconstruction error when they are combined. Therefore, all parameters need to be refined via a joint optimization. Given the initialized parameters $\Theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, since the output of CDA is $\hat{\mathbf{x}}_i$, the fine-tuning objective function is similar to that in (12), that is

$$\ell = \sum_i \left\|\mathbf{x}_i - \hat{\mathbf{x}}_i\right\|^2 \tag{15}$$

where $\mathbf{x}_i$ is the true value of the HR patch. The involved values $\hat{\mathbf{x}}_i$, $\mathbf{h}_i^H$, and $\mathbf{h}_i^L$ follow (4), (7), and (8), respectively. As with most deep learning methods, we use back-propagation to optimize (15).

The training steps of CDA are summarized in Algorithm 2.

Algorithm 2 Training of CDA

Input: LR/HR image patch sets $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ and $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in which each subscript indicates a LR/HR image patch pair

Output: CDA with $\Theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$

Step 1: Initialize $\mathbf{W}_1$ and $\mathbf{b}_1$ by learning intrinsic representations of low-resolution image patches $\mathbf{Y}$;

Step 2: Initialize $\mathbf{W}_3$ and $\mathbf{b}_3$ by learning intrinsic representations of low-resolution image patches $\mathbf{X}$;

Step 3: Initialize $\mathbf{W}_2$ and $\mathbf{b}_2$ by learning the nonlinear mapping from the LR intrinsic representations to the HR intrinsic representations;

Step 4: Construct CDA with the parameters $\Theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ and fine-tune them using standard back-propagation.

D. Discussion With Respect to Sparse Coding-Based SR

In this section, we discuss the advantages of the proposed CDA over SC for single image SR.

Generally speaking, in SC-based methods, a local patch extracted from the input LR image is subtracted by its mean and projected onto an LR dictionary. If the dictionary size is n_1, SC is similar to the autoencoder used in CDA for extracting an n_1-dimensional intrinsic representation. This is illustrated as stage 1 in Fig. 2. Taking the reconstruction into consideration, SC synthesizes the HR image patch based on the estimated HR sparse codes and an HR dictionary. This is similar to applying the autoencoder (see stage 2 in Fig. 2) to extract the n_2-dimensional intrinsic representation when the size of the HR dictionary is n_2. Usually, $n_1 = n_2$, in SC. Once the LR and HR sparse coefficients are obtained, various mapping techniques, such as identical mapping [13], linear mapping [15], [32], manifold similar mapping [12], and same support set mapping [40], can be built according to their respective assumptions. Of note, each of these mappings can be viewed as a special case of the mapping learned by CDA. This is because CDA does not rely on any assumptions, and the true mapping from the LR representation to the corresponding HR representation is discovered from extensive data.

CDA-based SR can therefore be regarded as a generalization of the conventional SC-based SR. CDA provides more flexible representations and better learning than SC because not all operations are considered during optimization in SC-based SR. In contrast, LR encoding, HR decoding, and nonlinear mapping are all involved in the deep learning model and are jointly optimized in CDA. In this way, our method optimizes an end-to-end mapping [48] consisting of all the operations.

III. Experiments

To demonstrate the effectiveness and efficiency of CDA, we conducted experiments on the same training and testing sets as those in [50]. Specifically, 91 images were used for training. For testing, Set5 [49] (five images) was used to assess the performance of upscaling factors 2–4, and Set14 [51] (14 images) was used to evaluate the upscaling factor 3. To objectively assess the quality of SR reconstruction, peak signal to noise ratio (PSNR) [52] were employed.

TABLE I
Results of PSNR (dB) on the Dataset of Set5. Without Pretraining, the Average PSNR of SR is 0.86 dB, 0.30 dB, and 0.27 dB Lower Than Those With Pretraining Upscaling the Images in Set5 When Upscaling Factor is 2, 3, and 4, Respectively

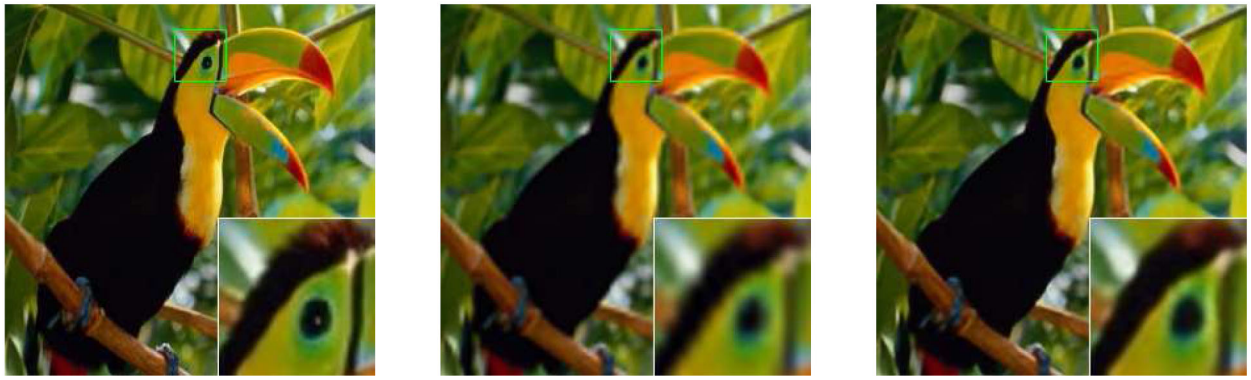
Set5 images	scale	CDA without pretraining	CDA with pretraining
baby	2	38.18	**38.29**
bird	2	39.20	**40.44**
butterfly	2	30.59	**32.50**
head	2	35.44	**35.55**
woman	2	34.25	**35.19**
average	2	35.53	**36.39**
baby	3	35.17	**35.23**
bird	3	34.66	**35.31**
butterfly	3	27.10	**27.84**
head	3	33.50	**33.61**
woman	3	30.75	**31.21**
average	3	32.24	**32.64**
baby	4	33.18	**33.23**
bird	4	32.07	**32.36**
butterfly	4	24.72	**25.28**
head	4	32.20	**32.30**
woman	4	28.28	**28.61**
average	4	30.09	**30.36**

TABLE II
Results of PSNR (dB) on the Dataset of Set14. When Dealing With the Set14, the Average PSNR of Super-Resoluting via CDA Without Pretraining is 0.24 dB Lower Than CDA With Pretraining

Set14 images	scale	CDA without pretraining	CDA with pretraining
baboon	3	23.60	**23.66**
barbara	3	**26.82**	26.65
bridge	3	25.09	**25.14**
coastguard	3	27.20	**27.25**
comic	3	24.28	**24.53**
face	3	33.48	**33.61**
flowers	3	28.86	**29.13**
foreman	3	33.16	**33.74**
lenna	3	33.30	**33.58**
man	3	28.04	**28.29**
monarch	3	32.08	**32.70**
pepper	3	34.10	**34.57**
ppt3	3	25.99	**26.31**
zebra	3	29.01	**29.16**
average	3	28.93	**29.17**

A. Experimental Configurations

For experiments involving CDA, the following optimal parameters were determined empirically. A patch size of 6×6, 9×9, and 12×12 was adopted when the upscaling factor was 2, 3, and 4, respectively. Thus, the number of neurons in CDAs input layer was set to 36, 81, and 144 corresponding to upscaling factor 2, 3, and 4, respectively. CDAs hidden layers contain the same number of neurons set to 2.5 times that of the input layer. In our experiments, more neurons lead to a higher computational cost but did not significantly improve SR performance, while fewer neurons decreased SR performance. We also noted that the number of neurons in the hidden layers can be different, which is an advantage over



(a) (b) (c)

Fig. 3. Comparison of SR (×3) on *Bird* image. (a) Original *Bird* image. (b) SC. (c) CDA.

(a) (b) (c)

Fig. 4. Comparison of SR (×3) on *Woman* image. (a) Original *Woman* image. (b) BPJDL. (c) CDA.

the SC-based SR method in [13]. To handle color images, the luminance channel in the YCrCb color space was considered. Two chrominance channels were directly upscaled using the bi-cubic method for the purpose of comparison with other SR methods. In order to obtain even better quality HR images, our method can be applied to three channels rather than the luminance channel alone.

All experiments in this section were performed on a workstation installed with an Intel core i7 CPU and 6 GB of memory.

B. Training of CDA With and Without Pretraining

There are two parts, which are initialization part and fine-tuning part, in the training procedure of CDA. The initialization part is essential to the training of CDA. In order to show its significance, two CDAs are trained. One trained according to the Algorithm 2 is named CDA with pretraining. The other trained according to the Algorithm 2 without the initialization part is named CDA without pretraining. Their SR results are shown in Tables I and II. Without pretraining, the average PSNR of SR is 0.86 dB, 0.30 dB, and 0.27 dB lower than those with pretraining upscaling the images in Set5 when upscaling factor is 2, 3, and 4, respectively. When dealing with the Set14, the average PSNR of super-resoluting via CDA without pretraining is 0.24 dB lower.

It is proven that the pretraining to CDA, which is the key to the success of CDA for image SR. The parameter space of a deep neural network is extremely large, and an improper initialization of the model may result in a poor solution after training. The pretraining initializes parameters to render them as close as possible to the solution which captures the intrinsic statistical structure of the input data. The fine-tuned parameters preserve these statistics and thus improve the performance.

C. Comparison Baselines

Seven state-of-the-art SR methods were considered as baselines: 1) the conventional bi-cubic method (Bicubic); 2) SC for SR [13]; 3) beta process joint dictionary learning (BPJDL) for SR [20]; 4) NE with non-negative least squares (NNLSs) [49];

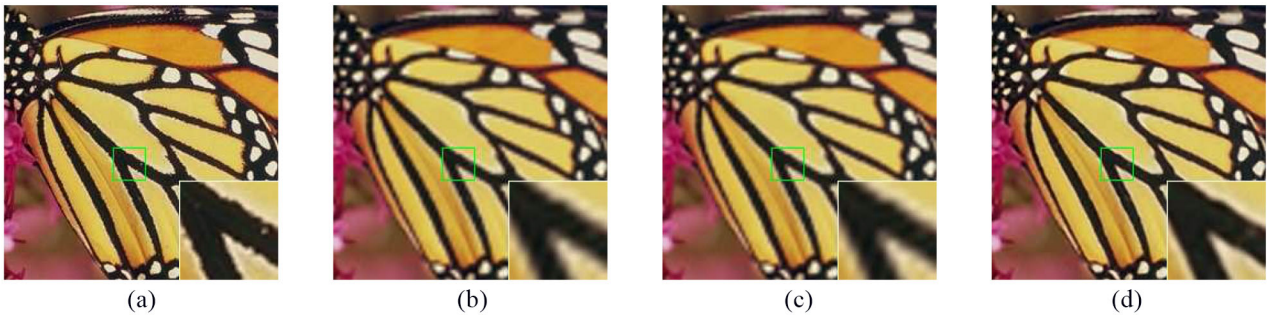


Fig. 5. Comparison of SR (×3) on *Butterfly* image. (a) Original *Butterfly* image. (b) NNLS. (c) LLE. (d) CDA.

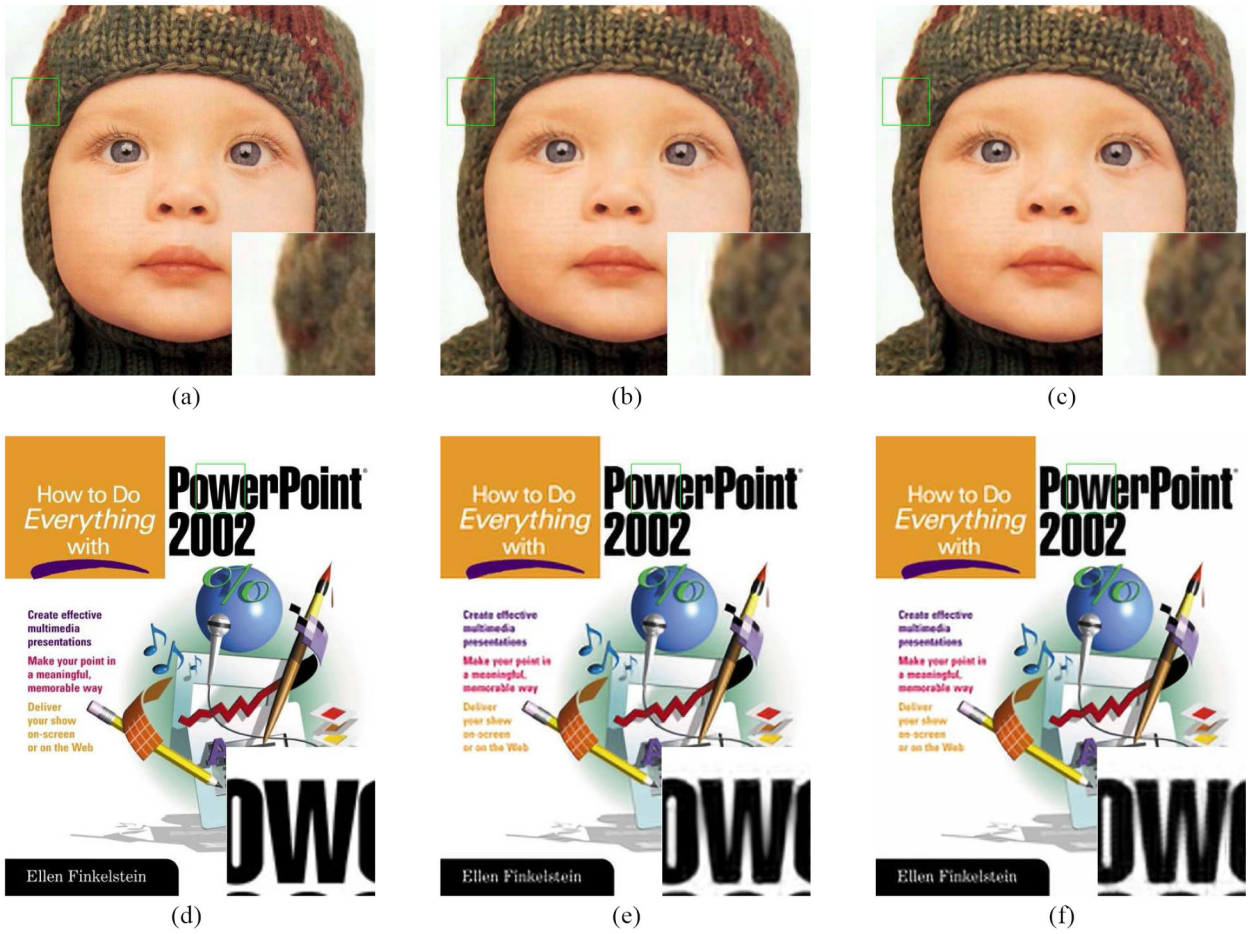


Fig. 6. Comparison of SR (×3) on *Baby* and *Ppt3* images. (a) and (d) Original *Baby* and *Ppt3* images. (b) and (e) CNN. (c) and (f) CDA.

5) NE with LLE [12]; 6) anchored neighborhood regression (ANR) [50]; and 7) SR via a convolutional neural network (CNN) [51]. The MATLAB implementations are all from the publicly available codes provided by the authors.

D. Performance Evaluations

As shown in Fig. 7 and Tables III and V, CDA significantly outperforms the other SR methods in all experiments and delivers the highest average PSNR, 0.27 dB higher than that of the second best approach, CNN, at best.

For subjective assessment, four sets of restored images using different methods are shown in Figs. 3–6. The SR results of SC and CDA on a *Bird* image are shown in Fig. 3; the SC-restored image is severely blurred and there are noticeable artifacts across the whole image, whereas CDA is sharper with suppression of edge artifacts. A *Woman* image is used to compare the performance of BPJDL and CDA in Fig. 4. The results indicate that BPJDL may bring blurring effects along the finger edges whereas CDA does not. SR results of two manifold mapping methods (NNLS and LLE) and CDA are shown in Fig. 5. Note the magnified plumage of the butterfly, where it is easy to observe jaggy artifact and blurring effects with NNLS and LLE. In comparison, CDA is much smoother, and, notably, the white stripe is clearly distinguished against the amber textures. In Fig. 6, CNN is similar to CDA, and there are few observable differences. However, the magnified subfigures

TABLE III
RESULTS OF PSNR (dB) ON THE DATASET OF SET5

Set5 images	scale	Bicubic	SC	BPJDL	NNLS	LLE	ANR	CNN	CDA
baby	2	37.07	-	**38.48**	38.00	38.33	38.44	38.30	38.29
bird	2	36.81	-	40.48	39.41	40.00	40.04	**40.64**	40.44
butterfly	2	27.43	-	31.00	30.03	30.38	30.48	32.20	**32.50**
head	2	34.86	-	**35.70**	35.48	35.63	35.66	35.64	35.55
woman	2	32.14	-	34.90	34.24	34.52	34.55	34.94	**35.19**
average	2	33.66	-	36.11	35.43	35.77	35.83	36.34	**36.39**
baby	3	33.91	34.39	**35.24**	34.77	35.06	35.13	35.01	35.23
bird	3	32.58	34.11	34.75	34.26	34.56	34.60	34.91	**35.31**
butterfly	3	24.04	25.58	26.45	25.61	25.75	25.90	27.58	**27.84**
head	3	32.88	33.17	**33.64**	33.45	33.60	**33.63**	33.55	33.61
woman	3	28.56	29.94	30.67	29.89	30.22	30.33	30.92	**31.21**
average	3	30.39	31.42	32.15	31.60	31.84	31.92	32.39	**32.64**
baby	4	31.78	-	33.09	32.81	32.99	33.03	32.98	**33.23**
bird	4	30.18	-	31.93	31.51	31.72	31.82	31.98	**32.36**
butterfly	4	22.10	-	23.89	23.30	23.28	23.52	25.07	**25.28**
head	4	31.59	-	32.20	32.10	32.24	32.27	32.19	**32.30**
woman	4	26.46	-	28.00	27.61	27.72	27.80	28.21	**28.61**
average	4	28.42	-	29.82	29.47	29.61	29.69	30.09	**30.36**

TABLE IV
RESULTS OF TIME (S) ON THE DATASET OF SET5

Set5 images	scale	Bicubic	SC	BPJDL	NNLS	LLE	ANR	CNN	CDA
baby	2	-	-	923.28	41.45	14.46	1.82	8.57	**0.92**
bird	2	-	-	251.45	13.03	4.48	0.59	1.66	**0.35**
butterfly	2	-	-	183.13	10.06	3.55	0.43	1.40	**0.26**
head	2	-	-	262.51	11.61	4.16	0.54	1.61	**0.31**
woman	2	-	-	243.63	11.81	4.16	0.51	1.49	**0.31**
average	2	-	-	372.80	17.59	6.16	0.78	2.95	**0.43**
baby	3	-	72.23	892.72	17.44	6.51	1.09	8.37	**0.80**
bird	3	-	29.24	256.68	5.34	1.97	0.34	1.67	**0.25**
butterfly	3	-	26.17	190.50	4.18	1.56	0.29	1.38	**0.20**
head	3	-	19.80	250.52	4.98	1.85	0.34	1.68	**0.24**
woman	3	-	23.23	238.82	5.00	1.84	0.32	1.61	**0.24**
average	3	-	34.34	365.85	7.39	2.75	0.48	2.94	**0.34**
baby	4	-	-	766.94	10.17	4.18	0.82	8.52	**0.81**
bird	4	-	-	228.83	3.01	1.25	**0.27**	1.70	0.31
butterfly	4	-	-	174.03	2.36	0.96	0.22	1.39	**0.20**
head	4	-	-	221.57	2.84	1.18	**0.24**	1.59	**0.24**
woman	4	-	-	213.18	2.81	1.15	0.27	1.66	**0.26**
average	4	-	-	320.91	4.24	1.75	0.36	2.97	**0.36**

TABLE V
RESULTS OF PSNR (dB) ON THE DATASET OF SET14

Set14 images	scale	Bicubic	SC	BPJDL	NNLS	LLE	ANR	CNN	CDA
baboon	3	23.21	23.47	23.56	23.49	23.55	23.56	23.60	**23.66**
barbara	3	26.25	26.39	**26.81**	26.67	26.74	26.69	26.66	26.65
bridge	3	24.40	24.82	25.09	24.86	24.98	25.01	25.07	**25.14**
coastguard	3	26.55	27.02	27.17	27.00	27.07	27.08	27.20	**27.25**
comic	3	23.12	23.90	24.19	23.83	23.98	24.04	24.39	**24.53**
face	3	32.82	33.11	33.58	33.45	33.56	**33.62**	33.58	33.61
flowers	3	27.23	28.25	28.70	28.21	28.38	28.49	28.97	**29.13**
foreman	3	31.17	32.04	33.13	32.87	33.21	33.23	33.35	**33.74**
lenna	3	31.67	32.64	33.26	32.82	33.01	33.08	33.39	**33.58**
man	3	27.01	27.76	28.07	27.72	27.87	27.92	28.18	**28.29**
monarch	3	29.43	30.71	31.55	30.76	30.95	31.09	32.39	**32.70**
pepper	3	32.38	33.33	34.05	33.56	33.80	33.82	34.35	**34.57**
ppt3	3	23.71	24.98	25.33	24.81	24.94	25.03	26.02	**26.31**
zebra	3	26.63	27.95	28.83	28.12	28.31	28.43	28.87	**29.16**
average	3	27.54	28.31	28.81	28.44	28.60	28.65	29.00	**29.17**

indicate that CNN produces ringing artifact along the edges of the hat in the *Baby* image and unwanted artifacts near the black edges of the letters in the *Ppt3* image. CDA handles both cases well and produces better detail. We can conclude that CDA offers better performance than the other learning-based methods.

TABLE VI
RESULTS OF TIME (S) ON THE DATASET OF SET14

Set14 images	scale	Bicubic	SC	BPJDL	NNLS	LLE	ANR	CNN	CDA
baboon	3	-	116.89	807.98	15.68	5.89	0.98	7.57	**0.73**
barbara	3	-	119.83	1354.03	27.87	10.44	1.65	15.45	**1.23**
bridge	3	-	140.74	895.53	17.35	6.52	1.09	8.46	**0.78**
coastguard	3	-	33.95	324.71	6.76	2.49	0.41	2.00	**0.34**
comic	3	-	48.86	292.52	5.77	2.14	0.36	1.83	**0.26**
face	3	-	19.56	247.09	4.91	1.84	0.33	1.61	**0.22**
flowers	3	-	72.34	596.47	11.81	4.42	0.71	5.32	**0.52**
foreman	3	-	25.52	312.67	6.63	2.45	0.42	2.02	**0.30**
lenna	3	-	66.16	860.23	17.13	6.46	1.06	8.66	**0.75**
man	3	-	101.19	875.33	17.23	6.47	1.04	8.49	**0.84**
monarch	3	-	107.47	1310.61	26.37	9.82	1.56	14.34	**1.15**
pepper	3	-	63.37	866.62	17.31	6.48	1.05	8.00	**0.78**
ppt3	3	-	92.88	926.80	21.85	8.46	1.40	11.64	**1.00**
zebra	3	-	110.07	770.87	15.13	5.68	0.94	6.83	**0.70**
average	3	-	79.92	745.82	15.13	5.68	0.93	7.30	**0.68**

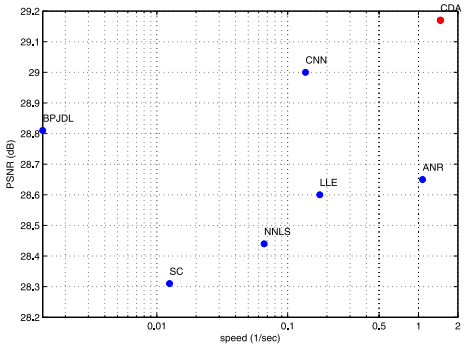


Fig. 7. Speed versus PSNR. CDA achieves state-of-the-art SR performance and is fastest compared to the other methods. The chart is based on the Set14 results summarized in Table V.

E. Running Time

Running time comparisons are shown in Tables IV and VI. CDA is clearly fastest when super-resolving LR images. Compared to SC, BPJDL, NNLS, and LLE, which need to solve complex optimization problems, ANR, CNN, and CDA are completely feed-forward. ANR maps the LR patches linearly onto the HR domain by multiplying a proper projection matrix according to the local neighborhoods; ANR is therefore faster than the other methods except CDA. CDA is potentially faster than CNN because of the compact HR/LR representations in the nonlinear mapping. In CDA, the mapping is a single-layer perceptron network, whereas CNN uses a convolutional layer whose input and output are a set of feature maps. Note that CNN does not use the pooling layers for reconstruction consideration, which causes a heavy computational load due to the exponentially increased feature maps. As shown in Table IV, the significant speed increase of CDA is best observed from the $3\times$ magnification results for Set5, where CDA achieves a 98.1-times speed advantage over SC, 21.1-times advantage over NNLS, 7.8-times advantage over LLE, 1.4-times advantage over ANR, and 8.4-times advantage over CNN.

Speed versus PSNR is plotted in Fig. 7. It can be seen that CDA outperforms the other methods with respect to both effectiveness and efficiency.

IV. CONCLUSION

This paper presents a novel data-driven model named CDA for single image SR. CDA automatically learns the intrinsic representations of LR/HR image patches and their relationship by exploring extensive data. Using an effective joint optimization, the intrinsic representations and the mapping function are co-dependently and effectively learned. CDA therefore represents an end-to-end learning scheme for single image SR. The effectiveness of CDA is thoroughly demonstrated by extensive experimentation. In terms of efficiency, CDA indeed behaves as a simple feed-forward network that only involves multiplication, addition, and activation operations. It is thus very fast to infer the high quality HR image from a given input LR image. Future work will include investigating the reliability of CDA in other vision applications and extending CDA to alternative basic networks such as the denoising autoencoder and CNN.

REFERENCES

[1] M. O. Camponez, E. O. T. Salles, and M. Sarcinelli-Filho, "Super-resolution image reconstruction using nonparametric Bayesian INLA approximation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3491–3501, Aug. 2012.
[2] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Image and video restorations via nonlocal kernel regression," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 1035–1046, Jun. 2013.
[3] X. Lu and X. Li, "Multiresolution imaging," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 149–160, Jan. 2014.
[4] B. Langmann, W. Wehihs, O. Loffeld, and K. Hartmann, "Development and investigation of a long-range time-of-flight and color imaging system," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1372–1382, Aug. 2014.
[5] Y. Tang and Y. Yuan, "Learning from errors in super-resolution," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2143–2154, Nov. 2014.
[6] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin, "Super resolution using edge prior and single image detail synthesis," in *Proc. CVPR*, San Francisco, CA, USA, 2010, pp. 2400–2407.
[7] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. CVPR*, Anchorage, AK, USA, 2008, pp. 1–8.
[8] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang, "Non-local kernel regression for image and video restoration," in *Proc. ECCV*, Crete, Greece, 2010, pp. 566–579.
[9] X. You, Q. Li, D. Tao, W. Ou, and M. Gong, "Local metric learning for exemplar-based object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1265–1276, Aug. 2014.
[10] M. Song, D. Tao, C. Chen, X. Li, and C. W. Chen, "Color to gray: Visual cue preservation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1537–1552, Sep. 2010.

[11] Y. Tang and Y. Yuan, "Image pair analysis with matrix-value operator," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2042–2050, Oct. 2015.
[12] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. CVPR*, vol. 1. Washington, DC, USA, 2004, pp. 275–282.
[13] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. CVPR*, Anchorage, AK, USA, 2008, pp. 1–8.
[14] J. Yang, Z. Lin, and S. Cohen, "Fast image super-resolution based on in-place example regression," in *Proc. CVPR*, Portland, OR, USA, 2013, pp. 1059–1066.
[15] C.-Y. Yang and M.-H. Yang, "Fast direct super-resolution by simple functions," in *Proc. ICCV*, Sydney, NSW, Australia, 2013, pp. 561–568.
[16] X. Lu, H. Yuan, P. Yan, Y. Yuan, and X. Li, "Geometry constrained sparse coding for single image super-resolution," in *Proc. CVPR*, Providence, RI, USA, 2012, pp. 1648–1655.
[17] K. Zhang, X. Gao, D. Tao, and X. Li, "Multi-scale dictionary for single image super-resolution," in *Proc. CVPR*, Providence, RI, USA, 2012, pp. 1114–1121.
[18] Y. Zhu, Y. Zhang, and A. L. Yuille, "Single image super-resolution using deformable patches," in *Proc. CVPR*, Columbus, OH, USA, 2014, pp. 2917–2924.
[19] G. Polatkan, M. Zhou, L. Carin, D. Blei, and I. Daubechies, "A Bayesian nonparametric approach to image super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 2, pp. 346–358, Feb. 2015.
[20] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *Proc. CVPR*, Portland, OR, USA, 2013, pp. 345–352.
[21] T. Michaeli and M. Irani, "Nonparametric blind super-resolution," in *Proc. ICCV*, Sydney, NSW, Australia, 2013, pp. 945–952.
[22] A. Adler, Y. Hel-Or, and M. Elad, "A shrinkage learning approach for single image super-resolution with overcomplete representations," in *Proc. ECCV*, Heraklion, Greece, 2010, pp. 622–635.
[23] N. Efrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin, "Accurate blur models vs. image priors in single image super-resolution," in *Proc. ICCV*, Sydney, NSW, Australia, 2013, pp. 2832–2839.
[24] C. Fernandez-Granda and E. J. Candes, "Super-resolution via transform-invariant group-sparse regularization," in *Proc. ICCV*, Sydney, NSW, Australia, 2013, pp. 3336–3343.
[25] P. Purkait, N. R. Pal, and B. Chanda, "A fuzzy-rule-based approach for single frame super resolution," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2277–2290, May 2014.
[26] 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.
[27] H. Zhang, Y. Zhang, H. Li, and T. S. Huang, "Generative Bayesian image super resolution with natural image prior," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4054–4067, Sep. 2012.
[28] O. Le Meur, M. Ebdelli, and C. Guillemot, "Hierarchical super-resolution-based inpainting," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3779–3790, Oct. 2013.
[29] Y. Xie *et al.*, "Discriminative object tracking via sparse representation and online dictionary learning," *IEEE Trans. Cybern.*, vol. 44, no. 4, pp. 539–553, Apr. 2014.
[30] J. Wang *et al.*, "Robust face recognition via adaptive sparse representation," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2368–2378, Dec. 2014.
[31] K. Li, J. Yang, and J. Jiang, "Nonrigid structure from motion via sparse representation," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1401–1413, Aug. 2015.
[32] S. Wang, D. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proc. CVPR*, Providence, RI, USA, 2012, pp. 2216–2223.
[33] 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.
[34] 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.
[35] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, Aug. 2012.
[36] Y. Romano, M. Protter, and M. Elad, "Single image interpolation via adaptive nonlocal sparsity-based modeling," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3085–3098, Jul. 2014.
[37] H. Su, Y. Wu, and J. Zhou, "Super-resolution without dense flow," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1782–1795, Apr. 2012.
[38] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
[39] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
[40] K. Jia, X. Wang, and X. Tang, "Image transformation based on learning dictionaries across image spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 367–380, Feb. 2013.
[41] C. Hou, F. Nie, X. Li, D. Yi, and Y. Wu, "Joint embedding learning and sparse regression: A framework for unsupervised feature selection," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 793–804, Jun. 2014.
[42] Z. Zhu *et al.*, "An adaptive hybrid pattern for noise-robust texture analysis," *Pattern Recognit.*, vol. 48, no. 8, pp. 2592–2608, 2015.
[43] B. Yoshua, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
[44] H. Zhou, G.-B. Huang, Z. Lin, H. Wang, and Y. C. Soh, "Stacked extreme learning machines," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 2013–2025, Sep. 2015.
[45] M. Song, D. Tao, X. Huang, C. Chen, and J. Bu, "Three-dimensional face reconstruction from a single image by a coupled RBF network," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2887–2897, May 2012.
[46] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. NIPS*, Vancouver, BC, Canada, 2006, pp. 153–160.
[47] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. ICML*, Helsinki, Finland, 2008, pp. 1096–1103.
[48] M. Song *et al.*, "Probabilistic exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 341–357, Jan. 2012.
[49] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. BMVC*, Surrey, BC, Canada, 2012, pp. 1–10.
[50] R. Timofte, V. De, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. ICCV*, Sydney, NSW, Australia, 2013, pp. 1920–1927.
[51] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. ECCV*, Zürich, Switzerland, 2014, pp. 184–199.
[52] 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.

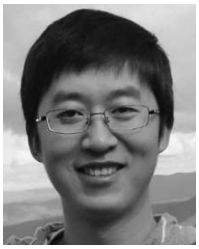
Kun Zeng received the B.Eng. and M.Sc. degrees from Computer Science Department, Xiamen University, Xiamen, China, in 2005 and 2008, respectively, where he is currently pursuing the Ph.D. degree with the School of Information Science and Technology.

His current research interests include image processing and machine learning.

Jun Yu (M'13) received the B.Eng. and Ph.D. degrees from Zhejiang University, Zhejiang, China.

He is currently a Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China. He was an Associate Professor with the School of Information Science and Technology, Xiamen University, Xiamen, China. From 2009 to 2011, he was with Singapore Nanyang Technological University, Singapore. From 2012 to 2013, he was a Visiting Researcher with Microsoft Research Asia, Beijing, China. He has authored and co-authored over 50 scientific articles. His current research interests include multimedia analysis, machine learning, and image processing.

Prof. Yu was the Co-Chair for several special sessions, invited sessions, and workshops. He served as a Program Committee Member or a Reviewer for top conferences and prestigious journals. He is a Professional Member of ACM and China Computer Federation.



Ruxin Wang received the B.Eng. degree from Xidian University, Xi'an, China, in 2010, and the M.Sc. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2013. He is currently pursuing the Ph.D. degree with the University of Technology Sydney, Ultimo, NSW, Australia.

His current research interests include deep learning and image processing.

Cuihua Li received the B.S. degree in computational mathematics from Shandong University, Jinan, China, in 1983, the M.S. degree in computational mathematics, and the Ph.D. degree in automatic control theory and engineering from Xi'an Jiaotong University, Xi'an, China, in 1989 and 1999, respectively.

He was an Associate Professor with the School of Science, Xi'an Jiaotong University before 1999. He is currently with the Department of Computer Science, Xiamen University, Xiamen, China. His current research interests include computer vision, video and image processing, and super-resolution image reconstruction algorithms.

Prof. Li is a member of the editorial boards of both the *Chinese Science Bulletin* and *Journal of Xiamen University Natural Science*.

Dacheng Tao (F'15) is a Professor of Computer Science with the Centre for Quantum Computation & Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia. He mainly applies statistics and mathematics to data analytics problems. His current research interests include 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 the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the *Journal of Machine Learning Research*, the *International Journal of Computer Vision*, the Conference on Neural Information Processing Systems, the International Conference on Machine Learning, the IEEE Conference on Computer Vision and Pattern Recognition, the IEEE International Conference on Computer Vision, the European Conference on Computer Vision, the International Conference on Artificial Intelligence and Statistics, the IEEE International Conference on Data Mining, and ACM ACM Knowledge Discovery and Data Mining.

Prof. Tao was a recipient of several best paper awards, including the Best Theory/Algorithm Paper Runner-Up Award in IEEE ICDM07, the Best Student Paper Award in IEEE ICDM13, and the 2014 ICDM 10 Year Highest-Impact Paper Award.