Retrieval Compensated Group Structured Sparsity for Image Super-Resolution

Jiaying Liu, *Member, IEEE,* Wenhan Yang, Xinfeng Zhang, and Zongming Guo, *Member, IEEE*

***Abstract*—Sparse representation-based image super-resolution is a well-studied topic; however, a general sparse framework that can utilize both internal and external dependencies remains unexplored. In this paper, we propose a group-structured sparse representation (GSSR) approach to make full use of both internal and external dependencies to facilitate image super-resolution. External compensated correlated information is introduced by a two-stage retrieval and refinement. First, in the global stage, the content-based features are exploited to select correlated external images. Then, in the local stage, the patch similarity, measured by the combination of content and high-frequency patch features is utilized to refine the selected external data. To better learn priors from the compensated external data based on the distribution of the internal data and further complement their advantages, nonlocal redundancy is incorporated into the sparse representation model to form a group sparsity framework based on an adaptive structured dictionary. Our proposed adaptive structured dictionary consists of two parts: one trained on internal data and the other trained on compensated external data. Both are organized in a cluster-based form. To provide the desired over-completeness property, when sparsely coding a given LR patch, the proposed structured dictionary is generated dynamically by combining several of the nearest internal and external orthogonal sub-dictionaries to the patch instead of selecting only the nearest one as in previous methods. Extensive experiments on image super-resolution validate the effectiveness and state-of-the-art performance of the proposed method. Additional experiments on contaminated and uncorrelated external data also demonstrate its superior robustness.**

***Index Terms*—Super-resolution, structured sparsity, internal method, external method, retrieval compensation.**

I. Introduction

IMAGE super-resolution (SR) aims to recover a high resolution (HR) image from one or more low resolution (LR) images. The quality degradations inherent to image acquisition, saving, and storage causes LR images to lose high frequency detail, which leads to image SR recovery being an ill-posed problem. To solve this problem, a *priori* knowledge is imposed. Thus, one important issue of image SR is to constrain SR recovery with proper priors.

Since 1984 [1], studies on image super-resolution have been investigated sequentially. Single image SR can be classified into three categories: interpolation-based, reconstruction-based and example learning-based. Interpolation-based methods [2, 3] utilize the correlation between pixels to construct a prediction function to estimate the missing pixels. Reconstruction-based methods adopt a maximum a *posteriori* probability (MAP) framework in which various regularization terms are imposed as prior knowledge to describe some desirable properties of natural images to constrain the solution of the ill-posed SR recovery problem. Typical regularization terms include gradient [4, 5], nonlocal [6–8] and total variation (TV) [9, 10]. For both interpolation-based and reconstruction-based methods, prior knowledge is typically achieved in a rather fixed or heuristic way. Thus, it is insufficient to represent the diversified patterns of natural images.

Example-based methods learn the mappings between LR and HR image patches from large training sets. Given an LR patch, its corresponding HR patch is estimated based on these learned mappings. In these methods, prior knowledge is dynamically learned rather than provided heuristically. Thus, the modeling capacity of example-based methods depends largely on the training data source. There are usually two kinds of training data sources: the LR data and external images, further dividing the example-based methods into two subclasses: internal and external SR methods.

Internal SR methods [7, 11–15] learn priors from a training set cropped from the LR image itself. Based on the self-similarity property (that some salient features repeat across different scales within an image), the coupled LR/HR patches extracted from a hierarchical pyramid of LR images provide an effective prior for building the inverse recovery mapping. In [14], a fast single image super-resolution method combines self-example learning and sparse representation by replacing the exact SVD and l_1 norm with K-SVD and l_0 norm to achieve rapid self-learning. In [7], nonlocal similarity, one important kind of self-similarity, is incorporated into the sparse representation model to constrain and improve the estimation of sparse coefficients. To add more diversified and abundant patterns to the internal dictionary, Huang *et al.* [16] proposed to expand the internal patch search space by localizing planes with detected perspective geometry variations in the LR image. In these methods, the patch priors are selected and learned from the LR images; thus they are good at reconstructing the repeated patterns in the LR image. However, the internal patch priors fail to cover the diversified patterns of natural images and are poor at reconstructing the distinct patterns. Moreover, the degraded LR image loses high-frequency details, limiting the modeling capacity of internal priors.

In contrast to the internal methods, external methods present complementary and desirable properties. These methods utilize

This work was supported by National High-tech Technology R&D Program (863 Program) of China under Grant 2014AA015205 and National Natural Science Foundation of China under contract No. 61472011.

The authors are with Institute of Computer Science and Technology, Peking University, Beijing 100871, China, e-mail: liujiaying@pku.edu.cn, yangwenhan@pku.edu.cn, xfzhang@ntu.edu.sg, guozongming@pku.edu.cn.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2016.2614427, IEEE Transactions on Multimedia

the general redundancy among natural images and learn the LR-HR mappings from large training datasets containing representative coupled external patches from an external dataset. Some external SR methods apply the learned priors to SR estimation directly, without any online auxiliary adaptation, thus they are categorized into *fixed external methods*, including neighbor embedding [17–19], kernel ridge regression [20], factor graph [21], kernel PCA [22], locality-constrained representation [23], coupled dictionary [24–27] and the recently proposed deep learning [28, 29]. Compared with the internal methods, when the training set containing a variety of reference images, the priors extracted are more representative and general. However, the fixed prior may not succeed in modeling some image patterns because of the limited numbers of model parameters and training images.

Another branch of methods - *adaptive external methods* adjust the learned prior based on the information in LR images, to make the external prior more adaptive. In [30], the patch prior is modeled as a flexible deformation flow rather than a fixed vector. These deformable patches are more similar to the given LR patch in the LR feature space. Thus, HR patches estimated based on the fusion of these deformable patches present more similar HR features. However, image degradation can make the LR information ambiguous; thus, the deformation estimated in the LR feature space may be imprecise. Rather than adjusting the dictionary or the training set to the LR image, some works perform online compensation, which selects and imports correlated external information to update the training set and models. In [31], an Internet-scale scene matching performs searches for ideal example textures to constrain image upsampling. In [32], with the help of a database containing HR/LR image segment pairs, high-resolution pixels are ”hallucinated” from their texturally similar segments. These two works focus on hallucinating visually pleasant texture regions in large-scale enlargements rather than on restoring the ground truth details. In [33], the semantic information from parsing is used to choose the corresponding anchor points adaptively to benefit anchor regression-based image SR. In [34], Yue *et al.* proposed a cloud-based landmark SR method that searches for similar patches in registered and aligned correlated images and utilizes these patches to compensate the lost HR details. In this method, the referenced correlated images play an important role in predicting the details lost in the degradation. When the correlated images are similar, such as adjacent frames of a video or images of the same landmark or object with slight viewpoint differences, the reconstruction is highly accurate. However, when the reference images are dissimilar, the performance of the reconstruction drops significantly.

Due to the obvious strengths and weaknesses of these two kinds of priors, as well as their strong complementary properties, recent works have attempted to utilize both internal and external priors for image denoising and image SR. In [35, 36], the advantages of internal and external denoising methods are measured; then, these two kinds of methods are combined by balancing the error between noise-fitting and signal-fitting. In [37], Burger *et al.* proposed a learning method to adaptively combine internal and external denoising results. Timofte *et al.* [38] explored seven ways to benefit image SR, one of which is to create an internal dictionary containing internal anchor points for further joint anchor regression with the external dictionary. Wang *et al.* [39] proposed a joint SR method to adaptively fuse the results of sparse coding for external examples and those of epitomic matching for internal examples. This fusion is implemented via an adaptive balance between the reconstruction performance based on the internal and external priors. However, the joint weighting fails to delve deeper into the interdependency of internal and external priors at the model level, such as organizing external data based on the structure of internal data. Thus, some complementary advantages of internal and external models are still unexplored. Moreover, the fixed external training set leads to an inconsistency between the distributions of internal and external data; thus, the method with only external priors may generate a biased reconstruction result. In essence, an ideal framework that makes full use of both internal and external data should fulfill four conditions:

- The interdependence between external and introduced internal data should be depicted, and the complementary properties of external and internal models should be characterized.
- The introduced external data should be adjusted based on the characteristics of the LR image to guarantee consistency between the distributions of the internal and external data.
- When introducing external data, the model should be robust to degradation and uncorrelated data.
- To make dynamically introduced external data convenient, the model should be trainable in real time, allowing adaptive retraining with updated external data.

Considering these properties, in this paper, we propose a group sparse representation model to introduce both internal and external data for image super-resolution. The contributions of our paper are as follows:

- To the best of our knowledge, this study is the first attempt to introduce, organize and exploit external data in a unified sparse representation model based on the content and the structure of internal LR data. Empirical evaluations demonstrate the effectiveness of our proposed method as well as its robustness to basic degradations and uncorrelated data.
- A group-structured sparse representation model with compensated external priors is proposed. The nonlocal redundancy is incorporated into the sparse representation model based on an over-complete dictionary generated dynamically from both introduced external data and internal data.
- A two-stage similarity refinement guarantees the similarity between the LR images and the introduced external information from searched images and further ensures the positive effect of imported external data on image SR.

The rest of this paper is organized as follows. Section II briefly reviews the sparse representation model. In Section III, we introduce the proposed two-stage similar patch retrieval approach to obtain refined external data. To utilize this useful external information effectively, in Section IV, we propose a group sparse coding based on an adaptive structured dictionary. Section V explores a method to exploit both the internal and searched external information to build an iterative integrated framework to super-resolve images based on the group structured sparse representation model. We evaluate the effectiveness of our proposed method through experiments in Section VI. Finally, concluding remarks are given in Section VII.

II. Overview of Sparse Representation

A. Sparse Coding

Sparse representation generalizes a signal transformation as a decomposition based on a limited subset of basis functions or signal atoms from a large over-complete dictionary. Formally, let $\mathbf{x}$ be an image patch and $\mathbf{\Phi}$ be an over-complete dictionary. Let α be the coefficient that represents $\mathbf{x}$ sparsely over $\mathbf{\Phi}$. The sparse representation model can be represented as

$$\arg\min_{\alpha} ||\mathbf{x} - \mathbf{\Phi}\alpha||_2^2 + \lambda||\alpha||_p, \tag{1}$$

where the first term is the data fidelity term and the second term is the sparsity prior. λ balances the importance of these two terms. The choice of p determines the properties of the solution. A value of $p = 0$ leads to an NP-hard problem solved by greedy pursuit algorithms such as orthogonal matching pursuit (OMP), whereas a value of $p = 1$ leads to a convex problem that can be solved by the basis pursuit (BP) [40] and the FOCal Underdetermined System Solver (FOCUSS) [41]. After α is acquired, the estimation of $\mathbf{x}$ is obtained as follows:

$$\hat{\mathbf{x}} = \mathbf{\Phi}\alpha. \tag{2}$$

The sparsity prior helps in extracting the principal components of image structures and in removing noises or insignificant details from images. Because of its intrinsic robustness, sparse representation is widely applied in various image restoration applications [24, 42, 43]. However, in the traditional sparse representation model, image patches are assumed to be independent and uncorrelated; therefore, spatial correlations of these patches are neglected.

Because natural images are highly structured, image patches and their corresponding representation coefficients are correlated. Similar patches in the spatial domain also present a strong correlation among their sparse coefficients. To model this type of structural property, structured sparse representation models [7, 8, 44] introduce context information (*i.e.*, the distribution of similar patches) to depict the correlation of dictionary atoms between patches, leading to a more effective model.

Nonlocal similarity reflects the fact that some salient structural regions such as edges and textures repeat within an image. The group sparsity prior [45] collects nonlocal similar patches into groups for sparse coding [7, 46]. It is usually formulated as follows:

$$\arg\min_{\alpha_g} ||\mathbf{x}_g - \mathbf{\Phi}\alpha_g||_F^2 + \lambda||\alpha_g||_{0,\infty}, \tag{3}$$

where $\mathbf{x}_g$ is the patch group, and α_g contains all the corresponding sparse coefficients of patches in a group. $||\cdot||_{0,\infty}$ denotes the number of nonzero rows in a matrix. By utilizing the strong correlations of representation coefficients, group sparsity pursues a stable and accurate sparse coding to mitigate the ambiguity caused by the degradation process. However, this framework is not computationally efficient due to the high complexity of solving the $||\cdot||_{0,\infty}$ regularized problem.

In our work, we are interested in incorporating the group sparsity prior into the sparse representation model to depict the statistical dependencies between dictionary atoms in the context of introducing both internal and external data. Our sparse representation model is built on the patch group and regularized by l_0 norm. Further, to improve the computational efficiency, it is solved by simultaneous orthogonal matching pursuit (SOMP) [47, 48] to sparsely code the given patch group over a subset of dictionary atoms rather than over the whole dictionary.

B. Dictionary Learning

In addition to sparse coding, dictionary learning is also a fundamental part of sparse representation. In general, dictionaries can be classified into several types: orthogonal dictionaries (also called analytical dictionaries) (DCT and wavelet), over-complete dictionaries [49] and structured dictionaries [7, 50, 51]. Orthogonal dictionaries consist of the basis for their corresponding transforms. Over-complete dictionaries [24, 26, 52–54] are learned based on a reconstructed performance of a fixed training set. An over-complete dictionary is modeled in the form of the sparse coding problem, but tries to jointly optimize the representation coefficients and the dictionary:

$$\arg\min_{\alpha,\mathbf{\Phi}} ||\mathbf{x} - \mathbf{\Phi}\alpha||_F^2 + \lambda||\alpha||_p. \tag{4}$$

Learned from natural images, these dictionaries increase their local adaptiveness and modeling capacity. They provide redundancy, a desirable property for image reconstruction. However, their training phases are time-consuming. Therefore, it is impractical to frequently update these dictionaries to facilitate real-time applications. Moreover, the structural regularity between atoms within the dictionary is ignored, and a universal dictionary may not be adaptive for modeling some local image regions.

Structured dictionaries [7, 50, 51] are constructed based on patch clusters. First, the training patches are clustered and then, sub-dictionaries are learned based on the patch clusters. Sparse decomposition on one patch is carried out with the corresponding sub-dictionary, making them highly adaptive to local structures. The process of training these dictionaries is highly efficient, which enables dynamic retraining using the LR image before the SR reconstruction. However, in these methods, sub-dictionaries are orthogonal, which limits their modeling capacity in describing complex natural image signals.

In our model, we try to combine the advantages of the over-complete and structured dictionaries by considering both the internal and external data. We train two structured dictionaries with internal and adaptively selected external training images and then use both dictionaries to jointly constrain the image reconstruction.

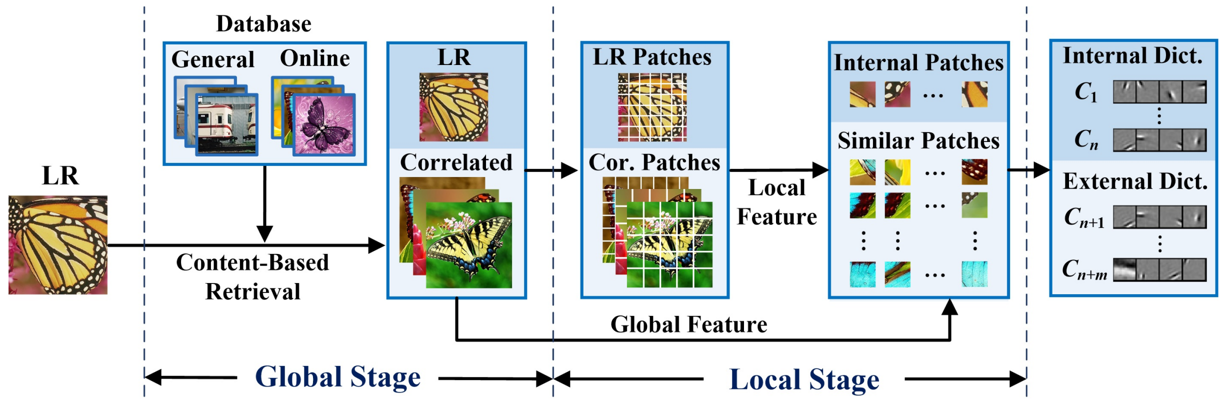


Fig. 1. Flow diagram of the introduced two-stage similarity-measured external information and subsequent dictionary learning based on the refined external similar patches and internal patches.

III. Two-Stage Similarity-Measured External Information Introduction

External data usually provides useful compensated information to reconstruct unrepeated patterns or structures of an image. However, at the same time, it may also introduce noise or irrelevant data that can degrade the subsequent reconstruction. Image degradation makes the LR-HR mappings ambiguous. Thus, selecting external information based on only LR images may be inaccurate. To avoid importing mismatched external data, we exploit both content and patch information to jointly select external information. Then, we propose a two-stage similarity measurement to refine external information as shown in Figure 2. In the global stage, the external information is first selected and refined based on content information. Then, in the local stage, the similarities between patches are measured jointly by content features and high frequency patch features.

A. Global-Stage Content-Based External Image Selection

Similar objects in some content consist of similar components. These objects tend to share similar low or middle level feature distributions such as skin colors of different people or texture patterns of beaches at different locations. Thus, the semantic and content information provides guidance when selecting useful external data. This intuition motivates us to use content-based image retrieval to search for correlated images and extract global features to facilitate the further patch matching. We first prepare an offline database containing various images. Then, before super-resolving an image, we use Google's search engine to obtain the first 10 similar images to an input LR image. These recalled images are added to the online database as a supplement to the offline database. We expect that this online database enhancement helps to simulate a cloud environment containing infinite images and ensures that our dataset always contains the images with similar content as the input LR. Note that Google's returned results contain both correlated and uncorrelated images; the correlated images provide useful information for further image selection and patch refinement.

For the retrieval, the features of Searching Images with MPEG-7-Powered Localized dEscriptors (SIMPLE) [55] is utilized. As with the majority of other recent popular features, SIMPLE combines the advantages of both global and local methods. It detects key points globally and forms features based on the corresponding patches locally. First, for a given image $\mathbf{X}$, a SURF detector [56] is used to detect key points $\{P_i\}$ in the image, where $i \in \{1, 2, \ldots, s\}$ and s is the number of the detected key points. Then, the local square region around P_i is defined as the salient image patch $\mathbf{L}_i$. Thus, the input image $\mathbf{X}$ is mapped into a series of salient image patches $\{\mathbf{L}_i\}$. Then, in each salient patch $\mathbf{L}_i$, a color and edge directivity descriptor (CEDD) $\mathbf{E}_i$ is extracted [57]. This is a 144-dimension vector that includes color, edge and texture information and has low complexity with high computational efficiency. Finally, the input image is represented by a set of CEDD features $\{\mathbf{E}_i\}$.

For indexing and retrieval, we use the BOW model [58]. All CEDD features $\{\mathbf{E}_i\}$ are quantified into visual words $\{\mathbf{E}_i^q\}$ through a local descriptor quantization. Then, we define the whole word set $\mathbf{W} = \{\mathbf{W}_j\}$, where $j \in \{1, 2, \ldots, t\}$, and t is the number of words. $\mathbf{W}$ contains all the quantized CEDD features. In the BOW model, an image is represented as the bag of its visual words $\mathbf{V} = \{(\mathbf{W}_j, n_j)\}$, where n_j is the number of the visual word $\mathbf{W}_j$ in the given image. Finally, the distance between two images $\mathbf{X}_u$ and $\mathbf{X}_v$ is defined as follows:

$$d\left(\mathbf{X}_u, \mathbf{X}_v\right) = \sum_{j=1}^{t} \left(n_j^u - n_j^v\right)^2, \tag{5}$$

where n_j^u and n_j^v are the numbers of the j-th visual word in the u-th and v-th image, respectively.

B. Local-Stage High-Frequency External Patch Matching

In the previous stage, we obtain images similar to the LR image. From these images, which are similar in both content and context, we further search for similar patches based on the concatenation of the global content feature and the high frequency patches. We split these similar images into patches $\{\mathbf{p}_n\}$, where $n \in \{1, 2, \ldots, N\}$, and N is the number of patches. Similar to the previous works [7, 24], the high frequency part of a patch $\mathbf{p}_n^h$ is estimated by the difference of Gaussians (DoG) operator. Then, a joint vector formed by concatenating the global stage content feature $\mathbf{V}_u$ and the local stage high frequency patch features $\mathbf{p}_n^h$ is used to represent a patch. The content feature of a patch is the SIMPLE feature of its corresponding image.

For indexing and retrieval, the KD-tree [59], an approximated nearest neighbor matching algorithm, is used. The joint vectors of all external patches are indexed. Then, the KD-tree searches for the best dimension in the feature space of the data to subdivide the reference dataset iteratively. This study used FLANN [60], an advanced KD-tree technique, to retrieve similar patches.

Overall, we utilize content similarity to shrink the selection range of external images in the global stage and search for similarly referenced patches based on both content and high frequency patch features in the local stage.

IV. Group Structured Sparse Representation

In this section, we construct a group structured sparse representation model that utilizes the nonlocal redundancy to constrain the sparse coding. Then, we illustrate our adaptive structured dictionary, including its composition and training algorithm. Note that the patches used for dictionary training consist of both the internal patch set $\mathbf{S}^I$ cropped from $\mathbf{Y}$ and the correlated external patch set $\mathbf{S}^E$ retrieved in Section III.

A. *Group-Based Sparse Representation*

Group-based sparse representation follows the patch representation framework. Let $\mathbf{X}$ be the HR image and $\mathbf{Y}$ be the LR image. $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ are overlapped patches cropped from $\mathbf{X}$ and $\mathbf{Y}$, where k indexes the locations of patches. For a single patch $\mathbf{x}_k$, a group of nonlocal similar patches $\mathbf{x}_k^g = \{\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \ldots, \mathbf{x}_{k,z}\}$ is collected based on the mean square error (MSE) between patches, where z is the number of similar patches in a group.

In general, the HR scene is degraded through blurring, down-sampling and noise addition operators to generate LR observations as follows:

$$\mathbf{y}_k = \mathbf{DH}\mathbf{x}_k + \mathbf{v}, \tag{6}$$

where $\mathbf{H}$ is the blur kernel, $\mathbf{D}$ is a down-sampling operator and $\mathbf{v}$ is the noise term. The traditional sparse representation models $\mathbf{x}_k$ as follows:

$$\mathbf{x}_k = \mathbf{\Phi}\alpha_k, \tag{7}$$

where α_k is the sparse coefficient that represents $\mathbf{x}_k$ over $\mathbf{\Phi}$. Then, the problem in (6) is converted to the problem of sparse coding for $\mathbf{y}_k$ with respect to $\mathbf{\Phi}$ as follows:

$$\hat{\alpha}_{\mathbf{x}} = \arg\min_{\alpha_k} \left\{ ||\mathbf{y}_k - \mathbf{DH\Phi}\alpha_k||_2^2 + \lambda||\alpha_k||_p \right\}. \tag{8}$$

The first term is the fidelity term, and the second term is the sparsity-inducing term. λ is a weighting parameter that makes a trade-off between the errors of these two terms.

Based on group sparsity, for a patch group $\mathbf{y}_k^g$ and $\mathbf{x}_k^g$, we construct a dynamic sub-dictionary $\mathbf{\Phi}_k$ (elaborated upon in Section IV-B) to represent $\mathbf{x}_k^g$. Meanwhile, we simplify the norm constraint from $l_{0,\infty}$ in (3) into l_0 norm. Then, (8) becomes the following problem of group sparse coding,

$$\begin{aligned} \hat{\alpha}_k^g &= \arg\min_{\alpha_k^g} \left\{ ||\mathbf{y}_k^g - \mathbf{HD\Phi}_k\alpha_k^g||_2^2 + \lambda||\alpha_k^g||_p \right\} \\ &= \arg\min_{\alpha_{k,m}} \left\{ \sum_{m=1}^{z} ||\mathbf{y}_{k,m} - \mathbf{HD\Phi}_k\alpha_{k,m}||_2^2 + \right. \\ &\qquad \left. \sum_{m=1}^{z} \lambda||\alpha_{k,m}||_p \right\}, \end{aligned} \tag{9}$$

where $\alpha_k^g = [\alpha_{k,1}, \alpha_{k,2}, and \ldots, \alpha_{k,z}]$ are the sparse representation coefficients of $\mathbf{x}_k^g$, and $\mathbf{\Phi}_k$ consists of a small subset of atoms dynamically selected from $\mathbf{\Phi}$ to represent $\mathbf{x}_k^g$. Equation (9), with the adaptive generated $\mathbf{\Phi}_k$, forces the nonlocal similar patches to have the same sparse decomposition pattern, and it can be solved by simultaneous orthogonal matching pursuit (SOMP) with $p = 0$. In our model, the group sparsity is in the form of the constraints in Section V. After obtaining the sparse coefficients, HR patches are reconstructed based on these coefficients and their corresponding dictionaries. Then, the entire image $\mathbf{X}$ is represented in the spatial domain by weighting the reconstructed patches.

B. *Adaptive Structured Dictionary Learning*

One important aspect of the sparse representation model is the dictionary $\mathbf{D}$. The analytical dictionaries such as DCT and wavelet are hand-crafted and orthogonal. The representation and reconstruction based on these dictionaries is equal to their corresponding transforms. They are compact but may fail to characterize some of the complex natural image signals. In contrast, learned dictionaries select a basis signal set to represent the image signal by measuring the reconstruction performance on a natural image training set. They are generally over-complete and their redundancy boosts the performance in depicting the complex image signals. However, their coding and reconstruction are usually related to the l_0 or l_1 optimization, which is unstable and time-consuming.

For our approach, we designed a stable and time-efficient over-complete dictionary—the adaptive structured dictionary—for sparse coding and image reconstruction in Section V. This approach forms an over-complete dictionary by combining several orthogonal sub-dictionaries that are trained based on the patch clusters sampled from a given image set. In sparse coding, several sub-dictionaries nearest to the given LR patch are chosen to form an over-complete dictionary. The entire process is shown in Fig. 2.

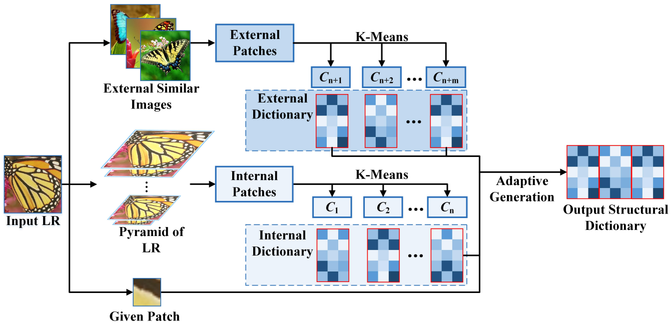


Fig. 2. The group-structured dictionary learning and online dictionary generation process in GSSR.

The patches used for the dictionary training are cropped from the LR image pyramid or from external images. Oversmooth patches are discarded using the condition $var(\mathbf{p}_i) < c$, where $var(\cdot)$ is the variance and c is the given threshold. Then, we acquire a training set $\mathbf{T} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M\}$ where M is the number of patches in $\mathbf{T}$. To obtain meaningful features, we extract the high-frequency versions of these patches $\mathbf{T}^h = \{\mathbf{p}_1^h, \mathbf{p}_2^h, \ldots, \mathbf{p}_M^h\}$ using the difference of Gaussians

(DoG) operator. The k-means algorithm is applied to divide $\mathbf{T}^h$ into K partitions $\{\mathbf{T}^h_1, \mathbf{T}^h_2, ..., \mathbf{T}^h_K\}$.

The traditional dictionary learning problem is modeled in the form of sparse coding in Eq. (8), which regards the dictionary as a variable to be estimated. However, it is time-consuming to directly solve it. For efficiency, we obtain the dictionary by applying the efficient PCA transformation to each patch cluster. For each cluster, let $\mathbf{\Omega}_k$ be the covariance matrix of the k-th partition $\mathbf{T}^h_k$. By applying the PCA to $\mathbf{\Omega}_k$, we get an orthogonal transform $\mathbf{F}_k$ in which the representation coefficients are $\mathbf{Z}_k = \mathbf{F}^T_k \mathbf{T}^h_k$. To make the model more compact and general, only parts of the eigenvectors are used to form $\mathbf{F}_k$. Thus, we limit the number of eigenvectors under a given threshold r. Let $\mathbf{F}_{k,r}$ and α_r be the transform matrix and representation coefficients with this limit, respectively. The proper r is then chosen as the optimal number of the eigenvectors involved in each cluster by solving the following optimization problem:

$$\hat{r} = arg \min_r \left\{ ||\mathbf{T}^h_k - \mathbf{F}_{k,r}\alpha_r||^2_F + \lambda||\alpha_r||_1 \right\}, \tag{10}$$

where $||\cdot||_F$ is the Frobenius norm.

To reconstruct a patch $\mathbf{x}_k$ or a group $\mathbf{x}^g_k$, we select several sub-dictionaries to obtain an adaptive over-complete dictionary. Let μ_i represent the centroid of patch cluster i, and let $\mathbf{x}^h_k/\mathbf{x}^h_{k,1}$ represent the high frequency parts of $\mathbf{x}_k/\mathbf{x}_{k,1}$. The sub-dictionaries are selected based on the distances between $\mathbf{x}^h_k/\mathbf{x}^h_{k,1}$ and μ_i. The distance $\mathbf{d}_{i_k}/\mathbf{d}_{i_{k,1}}$ is defined as follows:

$$\mathbf{d}_{i_k} = ||\mathbf{x}^h_k - \mu_i||_2 \text{ or } \mathbf{d}_{i_{k,1}} = ||\mathbf{x}^h_{k,1} - \mu_i||_2. \tag{11}$$

Those sub-dictionaries $\mathbf{\Phi}_i$ whose corresponding cluster $\mathbf{C}_i$ includes the smallest distances to $\mathbf{x}^h_k/\mathbf{x}^h_{k,1}$ are used to construct the over-complete dictionary by $\mathbf{\Phi}_o$:

$$\mathbf{\Phi}_o = [\mathbf{\Phi}_{k_1}, \mathbf{\Phi}_{k_2}, ..., \mathbf{\Phi}_{k_V}], \tag{12}$$

where k_j indicates that the center of the dictionary $\mathbf{\Phi}_{k_j}$ is the j-th closest to $\mathbf{x}_k/\mathbf{x}_{k_1}$, and V is the number of the sub-dictionary in forming the adaptive structured dictionary.

V. Super-Resolution based on GSSR with Internal and External Data

In this section, we utilize both internal and searched external data to build an integrated iterative framework to super-resolve images. Our framework is based on an optimization function with two parts: the fused patch priors and the sparsity constraint. Given a patch $\mathbf{x}_k$, the entire optimization is as follows:

$$\arg\min_{\mathbf{x}_k, \alpha^g_k} E_{patch}(\mathbf{x}_k) + \lambda_0 E_{sparse}(\mathbf{x}_k, \alpha^g_k). \tag{13}$$

The first term exploits the internal nonlocal similar patches and external HR patches to form a constraint on the estimation of $\mathbf{x}_k$. The second term incorporates both external and internal dictionary priors into the optimization function. Finally, λ_0 balances the importance of these two terms in the reconstruction.

A. Fused Patch Estimation with Nonlocal Mean and External Coupled Patches

The image degradation leads to a loss of high frequency details. We aim to build a simple inference approach from LR space to HR space. For the internal method, the nonlocal mean (NLM) is an effective tool for mapping the LR patches to the corresponding HR patches. By assuming that the patterns of image patches are usually non-locally correlated, NLM methods obtain a better HR image estimation by replacing every pixel with a weighted average of its neighborhood. For the external method, many techniques such as kernel regression [61] or neighbor embedding [17] can be used. These techniques utilize the external coupled patches to build the mapping from LR patches to HR patches. For simplicity, we use a generalized NLM to acquire a high-frequency detail estimation with the internal nonlocal patches and the general external coupled patches. Therefore, $E_{patch}(\cdot)$ is designed as a combination of the internal NLM and the external generalized NLM:

$$\begin{aligned} E_{patch}(\mathbf{x}_k) &= ||\mathbf{x}_k - \sum_i w^I_{k,i}\mathbf{x}^I_{k,i}||^2_2 \\ &\quad +||\mathcal{H}\mathbf{x}_k - \sum_j w^E_{k,j}\mathcal{H}\mathbf{x}^E_{k,j}||^2_2, \\ w^I_{k,i} &= \frac{1}{W_1}\exp\left\{-||\mathbf{x}_k - \mathbf{x}^I_{k,i}||^2_2/h_1\right\}, \\ w^E_{k,j} &= \frac{1}{W_2}\exp\left\{-||\mathbf{x}_k - \mathbf{x}^E_{k,j}||^2_2/h_2\right\}, \end{aligned} \tag{14}$$

where $\left\{\mathbf{x}^I_{k,i}\right\}$ and $\left\{\mathbf{x}^E_{k,j}\right\}$ are similar patches searched and extracted from the pyramids of the LR image and external images, respectively, where $i \in \{1, 2, ..., t_1\}$ and $j \in \{1, 2, ..., t_2\}$. Here, t_1 and t_2 are the numbers of the internal similar patches and external similar patches, respectively. $\mathcal{H}$ is the high-pass filter that extracts the high frequency part from a given patch. The values $W_1 = \sum^{t_1}_{i=1} \exp\left\{-||\mathbf{x}_k - \mathbf{x}^I_{k,i}||^2_2/h_1\right\}$ and $W_2 = \sum^{t_2}_{j=1} \exp\left\{-||\mathbf{x}_k - \mathbf{x}^E_{k,j}||^2_2/h_2\right\}$ are normalization factors, while h_1 and h_2 are pre-determined scalars. Intuitively, the reconstructed $\mathbf{x}_k$ is expected to be close to the combination of $\sum_i w^I_{k,i}\mathbf{x}^I_{k,i}$ and $(\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T \sum_j w^E_{k,j}\mathcal{H}\mathbf{x}^E_{k,j}$. In (14), the first term is used to generate more salient repeated features in the image and to preserve the general geometric properties of natural images. The second term imports more abundant high frequency details from external patches.

B. Sparsity Constraint with Internal and External Dictionaries

When introducing more high frequency detail, some noise and uncorrelated data may be introduced as well, causing additional visual degradation. To depress these artifacts and preserve the intrinsic geometric structures, we incorporate the GSSR constraint described in Section IV into the optimization function. For the dictionary prior, we use a concatenation of the internal and external dictionaries as the dictionary. The combination both strengthens their advantages and inhibits their individual weaknesses. This process can be written as follows:

$$\mathbf{\Phi}_k = \left[\mathbf{\Phi}_{E,k}, \mathbf{\Phi}_{I,k}\right], \tag{15}$$

where $\mathbf{\Phi}_{E,k}$ and $\mathbf{\Phi}_{I,k}$ are the external and internal dictionaries, respectively. Then, the corresponding group spare coefficients can be represented as

$$\alpha^g_k = \left[\alpha^g_{E,k}, \alpha^g_{I,k}\right]^T. \tag{16}$$

Then, (9) becomes

$$\begin{aligned} E_{sparse}\left(\mathbf{x}_k, \alpha_k^g\right) &= ||\mathbf{y}_k^g - \mathbf{HD}(\mathbf{\Phi}_k \alpha_k^g)||_F^2 + \lambda||\alpha_k^g||_0 \quad (17) \\ &= ||\mathbf{y}_k^g - \mathbf{HD}(\mathbf{\Phi}_{E,k}\alpha_{E,k}^g + \mathbf{\Phi}_{I,k}\alpha_{I,k}^g)||_F^2 \\ &\quad +\lambda||\alpha_{E,k}^g||_0 + \lambda||\alpha_{I,k}^g||_0. \end{aligned}$$

To adjust the preference given to the internal and external priors, we split the λ into two separated parameters: λ_1, λ_2. Then, (17) becomes

$$\begin{aligned} E_{sparse}(\mathbf{x}_k, \alpha_k^g) = \quad & ||\mathbf{y}_k^g - \mathbf{HD}(\mathbf{\Phi}_{E,k}\alpha_{E,k}^g + \mathbf{\Phi}_{I,k}\alpha_{I,k}^g)||_F^2 \\ & +\lambda_2||\alpha_{E,k}^g||_0 + \lambda_1||\alpha_{I,k}^g||_0. \end{aligned} \tag{18}$$

Because the internal dictionary is considered to be good at reconstructing some salient repeated patches within the given image, we give priority to the internal dictionary. Thus, we can rewrite (18) as follows:

$$\begin{aligned} & E_{sparse}(\mathbf{x}_k, \alpha_k^g) = \left\{||\mathbf{y}_k^g - \mathbf{HD\Phi}_{I,k}\alpha_{I,k}^g||_F^2 + \lambda_1||\alpha_{I,k}^g||_0 \right. \\ & \left. + \,||(\mathbf{y}_k^g - \mathbf{HD\Phi}_{I,k}\alpha_{I,k}^g) - \mathbf{HD\Phi}_{E,k}\alpha_{E,k}^g||_F^2 + \lambda_2||\alpha_{E,k}^g||_0\right\}. \end{aligned} \tag{19}$$

This reconstructs $\mathbf{y}_k^g$ with the internal dictionary first. Then, the external dictionary is utilized to rebuild the residual part, which is considered as a general pattern and therefore cannot be characterized by the internal dictionary prior.

C. Algorithm

We propose an alternating minimization method to solve (13). We split (13), turning it into several sub-problems by considering some variables as cyclically fixed.

1) α_I^g Problem: By fixing $\mathbf{x}_k$ and α_E^g, we obtain the following minimization problem:

$$\arg\min_{\alpha_{I,k}^g} ||\mathbf{y}_k^g - \mathbf{HD\Phi}_{I,k}\alpha_{I,k}^g||_F^2 + \lambda_1||\alpha_{I,k}^g||_0. \tag{20}$$

This is a problem of simultaneous orthogonal matching pursuit (SOMP) [47, 48]. When the group structure is fixed and the norm $||\cdot||_{0,\infty}$ is converted to the norm $||\cdot||_0$, we can solve it using SPAMS [62] software.

2) α_E^g Problem: By fixing $\mathbf{x}_k$ and α_I^g, we obtain a sub-problem concerning α_E^g:

$$\arg\min_{\alpha_{E,k}^g} ||(\mathbf{y}_k^g - \mathbf{HD\Phi}_{I,k}\alpha_{I,k}^g) - \mathbf{HD\Phi}_{E,k}\alpha_{E,k}^g||_F^2 + \lambda_1||\alpha_{E,k}^g||_0. \tag{21}$$

The problem in (21) can be solved in a similar way as (20).

3) $\mathbf{x}$ Problem: Finally, with α^g fixed, $\mathbf{x}_k$ can be solved simply as a weighted least squares (WLS) problem:

$$\begin{aligned} & \arg\min_{\mathbf{x}_k} ||\mathbf{x}_k - \sum_j w_{k,j}^I \mathbf{x}_{k,j}^I||_2^2 + ||\mathcal{H}\mathbf{x}_k - \sum_j w_{k,j}^E \mathcal{H}\mathbf{x}_{k,j}^E||_2^2 \\ & + \lambda_0||\mathbf{x}_k - \mathbf{\Phi}_{I,k}\alpha_{I,k}^1 - \mathbf{\Phi}_{E,k}\alpha_{E,k}^1||_2^2, \end{aligned} \tag{22}$$

where $\alpha_{I,k}^1$ and $\alpha_{E,k}^1$ are the sparse coefficients of the first patch in the patch group equal to the current patch. Let $\mathbf{w}_k^I = [w_{k,1}^I, w_{k,2}^I, ..., w_{k,t_1}^I]$, $\mathbf{x}_k^I = [\mathbf{x}_{k,1}^I, \mathbf{x}_{k,2}^I, ..., \mathbf{x}_{k,t_1}^I]$, $\mathbf{w}_k^E = [w_{k,1}^E, w_{k,2}^E, ..., w_{k,t_2}^E]$, $\mathbf{x}_k^E = [\mathbf{x}_{k,1}^E, \mathbf{x}_{k,2}^E, ..., \mathbf{x}_{k,t_2}^E]$. The problem in (22) can be reduced to:

$$\begin{aligned} & \arg\min_{\mathbf{x}_k} ||\mathbf{x}_k - \mathbf{w}_k^I(\mathbf{x}_k^I)^T||_2^2 + ||\mathcal{H}\mathbf{x}_k - \mathbf{w}_k^E\mathcal{H}(\mathbf{x}_k^E)^T||_2^2 \\ & + \lambda_0||\mathbf{x}_k - \mathbf{\Phi}_{I,k}\alpha_{I,k}^1 - \mathbf{\Phi}_{E,k}\alpha_{E,k}^1||_2^2 \end{aligned} , \tag{23}$$

which has a closed form solution:

$$\begin{aligned} & \mathbf{x}_k = \left[(\lambda_0 + 1)\mathbf{I} + \mathcal{H}^T\mathcal{H}\right]^{-1} \cdot \\ & \left[\mathbf{w}_k^I(\mathbf{x}_k^I)^T + \mathcal{H}^T\mathbf{w}_k^E\mathcal{H}(\mathbf{x}_k^E)^T + \lambda_0\mathbf{\Phi}_{I,k}\alpha_{I,k}^1 + \lambda_0\mathbf{\Phi}_{E,k}\alpha_{E,k}^1\right]. \end{aligned} \tag{24}$$

Then, the estimated HR image $\mathbf{X}$ is reconstructed by:

$$\hat{\mathbf{x}} = \left(\sum_{k=1}^{l}\mathbf{R}_k^T\mathbf{R}_k\right)^{-1}\sum_{k=1}^{l}\mathbf{R}_k^T\mathbf{x}_k. \tag{25}$$

D. Method Summary

We can summarize the solution of our model as an integrated, iterative SR framework (Fig. 3) that consists of two stages: patch enhancement and group sparse reconstruction. In the patch enhancement stage, both internal and external similar patches are fused to generate an HR estimation. Then, the given estimation is reconstructed by group sparse reconstruction based on both internal and external dictionary priors.

VI. Experimental Results

A. Experimental Setting

We built an image database by combining the PASCAL VOC Challenge 2012 dataset and web images retrieved from the Internet. The retrieved images come from the results returned by the Google image search engine in a manner similar to actual online updating. For each LR image, we added 100 similar images found on the web. Consequently, our database contains more than 19,000 images depicting various topics and a wide range of content types. In our tests, we selected the top five recalled images as the correlated images for each test LR image in the content image retrieval process. A typical example of the recalled images is shown in Fig. 4.

To verify the effectiveness of the introduction of the proposed two-stage similarity-measured external information and the group structured sparse representation for SR, we conducted extensive experiments on image enlargements. The basic parameter settings were as follows: 5 external similar patches were added to the training set for every LR patch; the patch size was 7 × 7; the overlap width was equal to 4; the number of clusters in the internal dictionary was $\mathbf{K}^I = 256$; the number of clusters in the external dictionary was $\mathbf{K}^E = 1024$; the group size was $z = 5$; and the adaptive generated over-complete dictionary contains 3 internal clusters and 3 external clusters. Other parameters are as follows: $h_1 = h_2 = 75$, $t_1 = t_2 = 10$, $\lambda_1 = \lambda_2 = 7$ and $\lambda_0 = 1$.

We conducted both qualitative and quantitative evaluations on our method, comparing it with the Bicubic interpolation method, ScSR [24], BPJDL [27], ASDS [63], NCSR [7], Landmark [64], SRCNN [28], ANR [18], A+ [19], SelfEx [16] and JSR [39]. The results of JSR and Landmark are provided by the author[1]. To make accurate comparisons possible, the source code for the compared methods was kindly provided by their authors. We followed the simulated image degradation process described in [7, 63], in which LR images are generated by a blurring and down-sampling operator. The blurring is performed with a 7 × 7 Gaussian kernel whose

[1] The input LR images were provided following the degradation process in this paper.

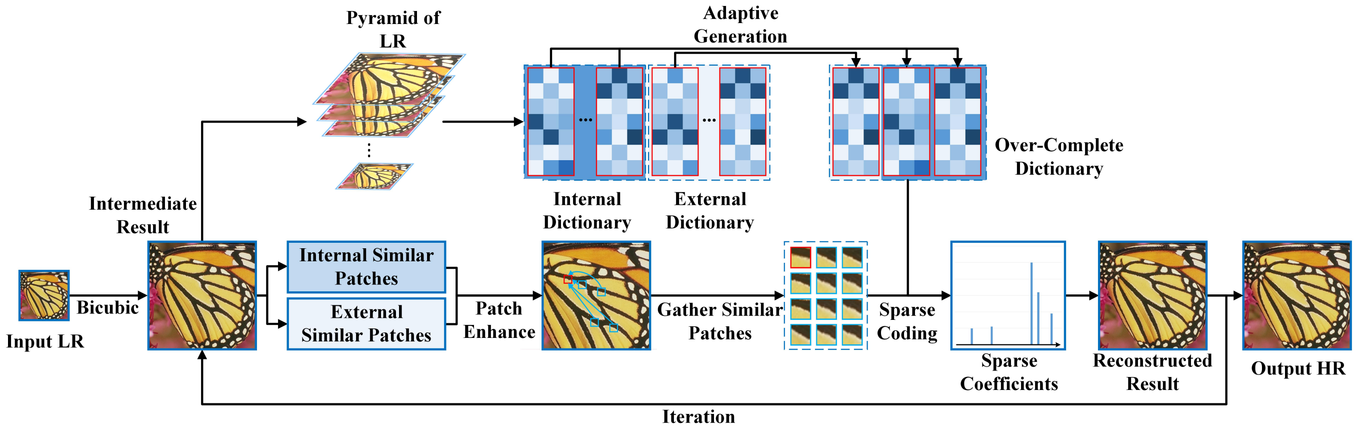


Fig. 3. A flow chart of the proposed super-resolution algorithm, including the iterative SR framework with patch enhancement based on both internal and external similar patches and the group sparse reconstruction based on the structured dictionary.

TABLE I
PSNR (dB) AND SSIM RESULTS IN 3× ENLARGEMENT.

Method	**Bicubic**		**ScSR**		**BPJDL**		**Landmark**		**SRCNN**		**SelfEX**	
Metric	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baboon	20.80	0.3547	21.50	0.4714	21.02	0.4234	21.02	0.4125	21.66	0.4924	21.10	0.4199
Bike	20.80	0.5759	23.36	0.7400	21.91	0.6680	21.68	0.6679	24.21	0.7736	21.02	0.6577
Butterfly	20.78	0.7175	25.14	0.8543	22.70	0.7887	21.92	0.7925	26.52	0.8664	22.64	0.7541
Car	24.66	0.7557	28.03	0.8595	26.17	0.8126	25.06	0.7902	28.64	0.8655	24.52	0.8036
Field	23.01	0.6248	24.57	0.7072	23.45	0.6662	23.25	0.6390	24.88	0.7134	23.95	0.6748
Comic	20.87	0.5573	23.53	0.7371	21.92	0.6607	21.59	0.6455	24.29	0.7684	21.58	0.6254
Foreman	26.48	0.8491	30.34	0.9151	28.07	0.8831	27.72	0.8820	29.87	0.9295	28.48	0.8666
Hat	27.20	0.7778	29.86	0.8449	28.10	0.8132	27.89	0.7946	30.50	0.8515	28.55	0.8267
Leaves	19.83	0.6411	24.40	0.8482	21.27	0.7523	19.21	0.6543	26.06	0.8754	21.37	0.7599
Lena	26.91	0.7660	30.73	0.8563	28.49	0.8133	27.46	0.7999	31.51	0.8675	28.70	0.8007
Text	10.80	0.4786	12.82	0.7000	11.66	0.6259	10.09	0.3840	13.44	0.7275	12.03	0.5913
Zebra	20.43	0.5398	24.14	0.7264	21.49	0.6390	20.88	0.6233	25.10	0.7566	21.04	0.7565
Average	21.88	0.6365	24.87	0.7717	23.02	0.7122	22.31	0.6738	25.56	0.7906	22.92	0.7114
Gain	*	*	2.99	0.1541	1.14	0.0757	0.43	0.0373	3.68	0.1541	1.03	0.0749
Method	**JSR**		**ANR**		**A+**		**ASDS**		**NCSR**		**Proposed**	
Metric	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baboon	20.96	0.4401	21.53	0.4684	21.61	0.4850	21.75	0.5042	21.75	**0.5068**	**21.76**	**0.5069**
Bike	20.70	0.6405	23.45	0.6593	24.18	0.6716	24.66	0.7983	24.72	0.8027	**25.02**	**0.8137**
Butterfly	20.74	0.7742	25.12	0.8547	26.44	0.8989	27.29	0.9034	28.08	0.9157	**28.87**	**0.9296**
Car	23.54	0.7596	28.11	0.8622	28.56	0.8826	29.36	0.8882	29.42	0.8915	**29.65**	**0.8928**
Field	21.45	0.6424	24.62	0.7091	25.05	0.7321	25.28	0.7368	25.40	**0.7413**	**25.56**	0.7390
Comic	20.46	0.6302	23.57	0.7374	24.02	0.7653	24.60	0.7869	24.65	0.7908	**24.85**	**0.8012**
Foreman	23.78	0.8767	30.74	0.9182	28.48	0.9330	31.72	0.9332	32.10	0.9358	**32.21**	**0.9381**
Hat	26.43	0.8100	29.88	0.8463	30.78	0.8671	30.97	0.8650	31.27	0.8705	**31.60**	**0.8775**
Leaves	19.17	0.7105	24.35	0.8458	25.32	0.8897	26.76	0.9066	27.43	0.9215	**28.26**	**0.9381**
Lena	25.68	0.7865	30.82	0.8603	31.59	0.8763	32.05	0.8806	32.25	**0.8844**	**32.27**	**0.8839**
Text	11.22	0.5527	12.79	0.6824	13.19	0.7264	11.55	0.5975	14.08	0.7718	**14.57**	**0.8119**
Zebra	21.07	0.6196	24.39	0.7310	24.97	0.7517	25.31	0.7656	25.52	0.7722	**25.80**	**0.7750**
Average	21.27	0.6869	24.95	0.7646	25.35	0.7900	25.94	0.7972	26.39	0.8171	**26.70**	**0.8256**
Gain	-0.61	0.05	3.07	0.1281	3.47	0.1534	4.06	0.1607	4.51	0.1806	4.82	0.1891

standard deviation is 1.6. Similar to previous works, the image SR methods are applied only to the luminance component, while the chromatic components are enlarged by the Bicubic interpolation. To evaluate the quality of the SR results, the Peak Signal-to-Noise Ratio (PSNR) and the perceptual quality metric Structural SIMilarity (SSIM) were calculated.

For SRCNN, ANR and A+, we retrained their network or dictionaries with our degradation setting and kept other configuration same as shown in original papers. The training set of SRCNN, created in [28], contained 91 images. They were cropped into 33×33 input and 21×21 output patches. These images were decomposed into around 15,000 sub-images using a stride of 21. SRCNN was trained on Caffe platform [65] via stochastic gradient descent (SGD) with standard backpropagation. We set the momentum as 0.9, the learning rate as a fixed value 10^{-4} for front-end layers and 10^{-5} for the penultimate layer during the training. We allowed at most 5×10^7 backpropagations, namely 2.2×10^5 epochs, which spent about three days on a single GPU – GTX 780Ti. We did not allow a larger number of backpropagations as reported in [28] because we did not observe further performance gain. For ANR and A+, we used the standard setting illustrated in [18] and [19] with 5 million training samples of LR and HR patches from the same training images in [28], a dictionary size of 1024, and a neighborhood size of 2048 training samples for A+ and 40 atoms for ANR, respectively. The training process of ANR and A+ cost about 10 to 15 minutes, which was much faster than SRCNN. ScSR, BPJDL, Landmark and SelfEx cannot perform enlargement and deblurring simultaneously; thus, an iterative back-projection was carried out for deblurring before the SR—the same as the preprocessing performed in [7]. The number of deconvolution iterations was set to obtain the best average PSNR result for each method.

B. Objective Evaluation

Table I and Table II lists the image SR results of our method and the five comparison methods using scaling factors

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2016.2614427, IEEE Transactions on Multimedia

TABLE II
PSNR (dB) and SSIM results in 2× enlargement.

Method	**Bicubic**		**ScSR**		**BPJDL**		**Landmark**		**SRCNN**		**SelfEX**	
Metric	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baboon	21.23	0.4036	21.92	0.5357	21.54	0.5118	21.91	0.5364	22.18	0.5598	21.07	0.5037
Bike	21.90	0.6478	24.15	0.7899	23.22	0.7645	24.38	0.8041	23.45	0.7246	23.84	0.7631
Butterfly	22.42	0.7802	26.13	0.8792	24.51	0.8568	25.60	0.8822	29.48	0.9231	24.48	0.8430
Car	26.30	0.8122	29.60	0.8979	27.84	0.8768	28.32	0.8865	31.85	0.9249	27.84	0.8661
Field	23.87	0.6661	25.27	0.7525	24.39	0.7316	25.13	0.7427	26.47	0.7829	24.89	0.7444
Comic	22.15	0.6415	24.60	0.8023	23.35	0.7691	24.58	0.8092	26.50	0.8551	23.86	0.7656
Foreman	28.21	0.8805	31.26	0.9256	29.33	0.9144	30.60	0.9219	32.32	0.9382	29.99	0.9120
Hat	28.30	0.8081	30.50	0.8648	29.19	0.8536	30.41	0.8599	32.30	0.8850	29.25	0.8644
Leaves	21.62	0.7378	25.84	0.8902	23.36	0.8518	22.44	0.8333	29.76	0.9438	23.05	0.8676
Lena	28.63	0.8121	31.46	0.8799	29.74	0.8600	30.69	0.8744	33.12	0.8991	29.28	0.8698
Text	11.74	0.5719	14.37	0.7917	13.38	0.7584	12.23	0.6261	15.64	0.8476	13.01	0.7680
Zebra	22.01	0.6288	25.70	0.7980	23.50	0.7564	24.50	0.7920	28.05	0.8404	23.97	0.7496
Average	23.20	0.6992	25.90	0.8173	24.44	0.7921	25.07	0.7974	27.59	0.8437	24.54	0.7931
Gain	*	*	2.70	0.1181	1.24	0.0929	1.87	0.0982	4.39	0.1445	1.35	0.0939
Method	**JSR**		**ANR**		**A+**		**ASDS**		**NCSR**		**Proposed**	
Metric	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Baboon	23.43	0.6927	21.97	0.5321	22.08	0.5414	22.46	0.5906	22.48	0.5894	**22.54**	**0.6020**
Bike	24.25	0.8306	24.75	0.8137	25.55	0.8400	27.03	0.8749	27.12	0.8781	**27.56**	**0.8884**
Butterfly	24.99	0.8935	27.02	0.9031	28.49	0.9323	29.62	0.9370	30.71	0.9480	**31.28**	**0.9523**
Car	25.79	0.8493	29.86	0.9091	30.35	0.9216	32.28	0.9359	32.47	**0.9390**	**32.78**	0.9384
Field	23.85	0.8239	25.56	0.7656	26.01	0.7821	27.02	0.8143	27.21	0.8149	**27.37**	**0.8205**
Comic	23.49	0.8071	25.11	0.8166	25.61	0.8352	27.24	0.8796	27.25	0.8806	**27.58**	**0.8910**
Foreman	25.66	0.9370	31.57	0.9411	32.12	0.9498	33.62	0.9463	**34.01**	**0.9510**	33.97	0.9489
Hat	28.86	0.8987	31.21	0.8807	32.01	0.8940	32.85	0.9010	33.05	0.9044	**33.41**	**0.9075**
Leaves	22.96	0.8822	26.63	0.9100	27.81	0.9375	30.24	0.9543	31.18	0.9635	**31.73**	**0.9674**
Lena	29.18	0.9008	32.15	0.8931	32.92	0.9041	33.74	0.9121	33.83	**0.9132**	**34.02**	0.9121
Text	14.38	0.7767	14.63	0.7952	14.99	0.8221	12.16	0.6035	19.70	**0.9488**	**20.13**	0.9428
Zebra	23.89	0.7822	26.22	0.8113	26.81	0.8238	28.40	0.8604	28.65	0.8637	**29.14**	**0.8706**
Average	24.23	0.8396	26.39	0.8310	27.06	0.8487	28.06	0.8508	28.97	0.8829	**29.29**	**0.8868**
Gain	1.03	0.1403	3.19	0.1317	3.86	0.1494	4.86	0.1516	5.77	0.1837	**6.09**	**0.1876**

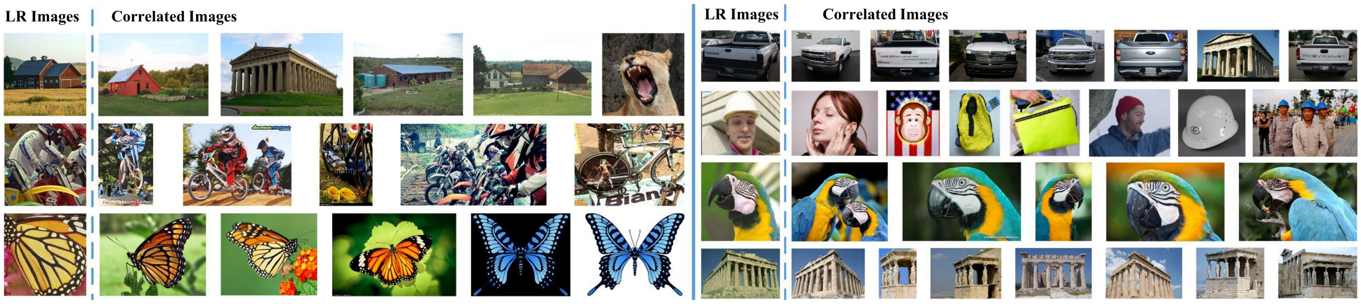
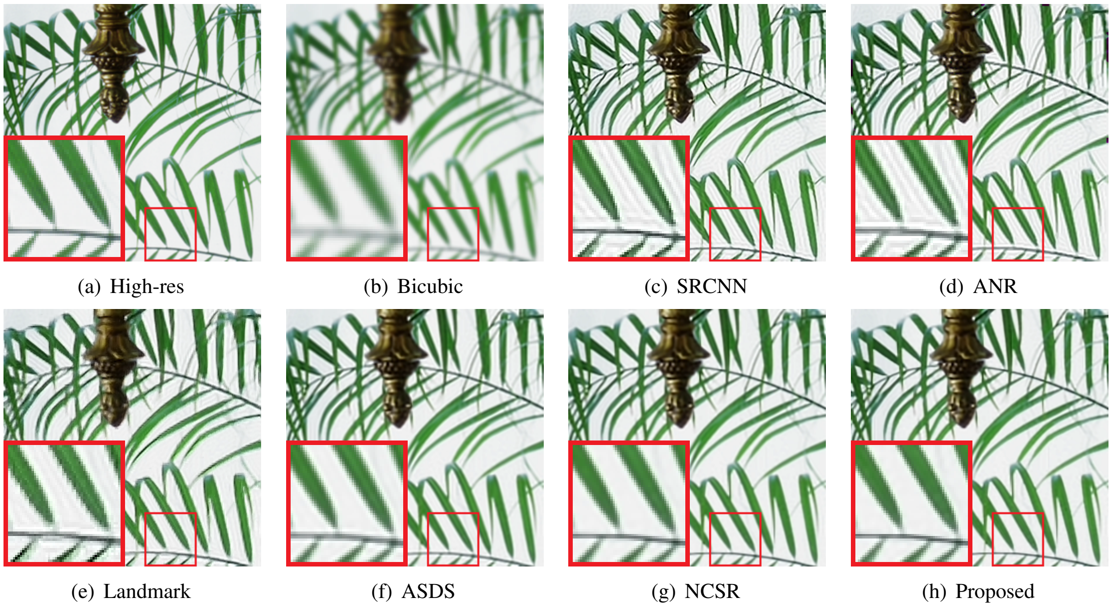


Fig. 4. Illustrations of the content images retrieved. For example, for *Car*, seven recalled images include only six similar images.

of 3 and 2, respectively. Our method outperformed the other SR methods for the majority of test images. In the 3× enlargement, our method achieved the best SR performance with averages of 26.68dB (PSNR) and 0.8252 (SSIM) over the 12 test images, constituting an improvement of 0.29dB in PSNR and 0.0081 in SSIM over the average results (26.39dB and 0.8171) of the second best method, NCSR [7]. Our method also achieved the best SR performance in the 2× enlargement with averages of 29.29dB (PSNR) and 0.8868 (SSIM). Here, the gain over NCSR is 0.32dB in PSNR and 0.0039 in SSIM. Four state-of-the-art methods, ANR, A+, SelfEx and SRCNN did not perform well in our experimental setting due to the heavy blurring. In a heavily blurred condition, the ambiguity between LR and HR spaces is enlarged and direct mapping methods such as similar patch fusion [64] and dictionary-based reconstruction [18, 24], degrade considerably.

C. Subjective Evaluation

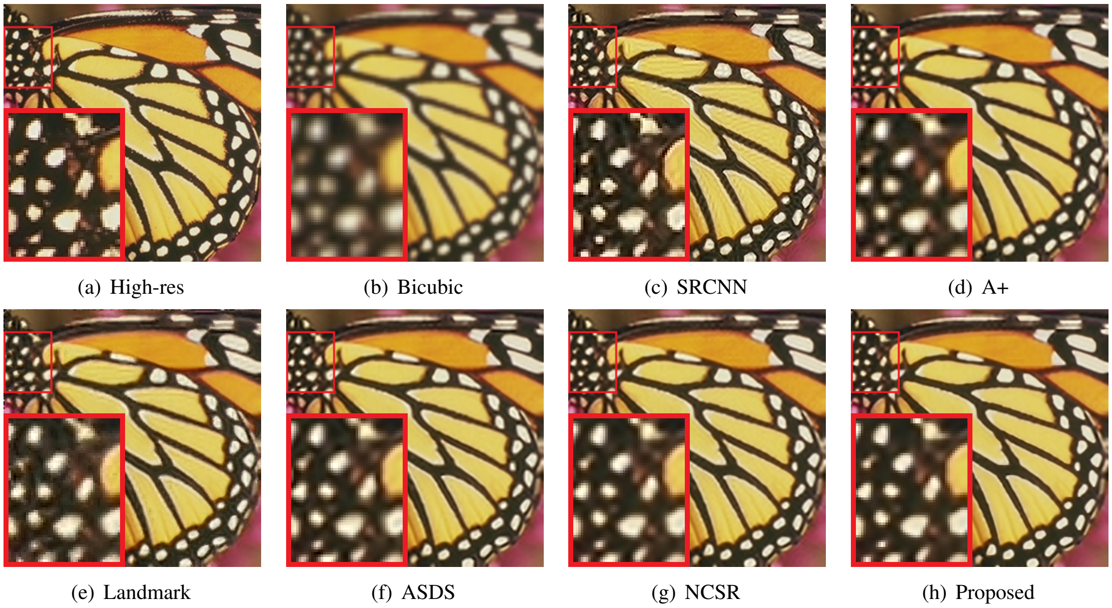
Fig. 5 demonstrates the super-resolution 2× results on *Leaves*. Fig. 6 shows the 3× results on *Butterfly*. As shown in these figures, the Bicubic interpolation generates blurred results. The ScSR method preserves the majority of edges, although there is still a little blurring around them. The ASDS method generates more natural edges and textures, but finds it difficult to avoid blurring and artifacts (*e.g.*, the stem in *Leaves*). Because the ASDS method is based on uncorrelated external images only, its sparse coding, which does not consider the consistency of the representation coefficients, is unstable. The NCSR recovers key structures such as the textures in *Butterfly*. However, it still introduces some blurring and slight (but noticeable) artifacts around the edges. This problem becomes more obvious for images with insufficient self-similarity. In the general case, the Landmark method cannot align the searched reference image with the input LR image, thus it degrades to a simple patch matching and fusion operation, leading to noise and artifacts in the reconstructed results. The SRCNN and ANR methods generate results containing obvious artifacts because they lack the ability to deal with the blurring and tend to enlarge the invisible artifacts generated by the deconvolution operator. In comparison, due to the combination of self-similarity and external similarity and by considering the cluster properties during dictionary training, our method preserves the edges better and generates more natural textures. **More subjective results are presented in the supplementary material.**



(a) High-res (b) Bicubic (c) SRCNN (d) ANR

(e) Landmark (f) ASDS (g) NCSR (h) Proposed

Fig. 5. Visual comparisons between different algorithms for the image *Leaves* (2×).



(a) High-res (b) Bicubic (c) SRCNN (d) A+

(e) Landmark (f) ASDS (g) NCSR (h) Proposed

Fig. 6. Visual comparisons between different algorithms for the image *Butterfly* (3×).

D. Ablation Analysis

To provide a closer look at the detailed performance of our method, we performed objective evaluations for several different versions of our method. We notate every version with an abbreviation. NCSR denotes the version that uses only internal information reconstructed by traditional sparse coding. **IEF** is based on both internal and fixed external data and reconstructed by traditional sparse coding. **IEC** is based on both internal and correlated external data and reconstructed by traditional sparse coding. **IECG** is based on both internal and correlated external data and reconstructed by group sparse coding. **ICGP** is based on only internal data and reconstructed by group sparse coding and fused patch priors. **IECGP** denotes a version based on the internal and correlated external data by group sparse coding and fused patch priors. Table III lists the PSNR results of our method in these different versions. The results show that the introduction of the external information, correlated external information, group sparsity constraints and fused patch priors improve the reconstruction performance in a step-by-step fashion. Moreover, the comparisons between the performances of ICGP, ECGP and IECGP indicate the strength of the combination of the external and internal data rather than the use of any single type.

E. Robustness of GSSR

To evaluate the robustness of our method, we tested its performance by introducing correlated but low-quality data, degrading the reference images used for training the external dictionaries with different noise levels. We also used a fixed, high-quality patch set to form the fused patch priors rather than sampling them from a noisy external patch set. We employed this strategy because we expect only the group sparse representation model to be robust to noises and regard the patch enhancement as being sensitive to image degradations. The results in Table IV indicate that the performance variance is

TABLE III
PSNR (dB) RESULTS OF DIFFERENT VERSIONS OF THE PROPOSED METHOD IN 3× ENLARGEMENT.

Method	NCSR	IEF	IEC	IECG	ICGP	ECGP	IECGP
Bike	24.72	24.82	24.83	24.83	24.79	24.06	**25.00**
Butterfly	28.08	28.34	28.47	28.52	28.50	26.57	**28.83**
Car	29.42	29.49	29.61	**29.67**	29.59	28.70	29.63
Comic	24.72	24.71	24.72	24.74	24.73	24.26	**24.85**
Field	25.40	25.55	**25.56**	25.54	**25.56**	25.08	**25.57**
Foreman	32.08	32.05	32.16	**32.24**	32.20	30.99	32.23
Hat	31.27	31.34	31.40	31.49	31.45	30.85	**31.59**
Leaves	27.43	27.76	27.87	27.84	27.81	24.64	**28.29**
Lena	32.25	32.28	32.32	32.37	**32.40**	31.61	32.34
Baboon	21.75	21.75	21.75	21.75	21.75	21.72	**21.75**
Text	14.08	**14.64**	14.25	14.31	14.26	13.48	14.52
Zebra	25.52	25.64	25.68	25.60	25.63	24.72	**25.80**
Average	26.39	26.53	26.55	26.58	26.56	25.56	**26.70**

negligibly small, meaning that the group sparse representation model is relatively insensitive to the quality of the introduced external images.

In fact, noisy external images seem to have almost no impact on the overall performance primarily for three reasons. First, when the external referenced images contain noise, the LR patch is far away from the centers of most external dictionaries/clusters, the sparse reconstruction of a patch tends to use internal sub-dictionaries to form the online dictionary instead of external sub-dictionaries. This explains why increasing the noise level in external images has little effect on the SR performance. Second, the external dictionaries trained from external referenced images with additive zero mean noises could at least contain some atoms that are able to describe high frequency details. Their existence enables the sparse representation using both internal and external dictionaries to preserve more structural details within a patch than an approach that uses only internal dictionaries. This explains why the external sub-dictionaries trained from noisy external referenced images still benefit the final SR result.

More surprisingly, it is observed from Table IV that, for one sample, *e. g. Baboon*, inputting random noise as the external images leads to a performance gain. We give a simple explanation here. The dictionary used in our paper aims to constrain the reconstruction to suppress the artifacts from other steps, instead of creating the correspondence between the LR and HR spaces. The original PCA dictionaries may be over-constrained and lead to removing irregular texture details. Then, when we relax the constraint to a certain extent even in a random way by providing sub-dictionaries learned from external noisy images, some irregular texture details, from internal nonlocal patches and external similar patches collected from a rather large fixed high-quality image pool, are better preserved.

F. Super-Resolving Noisy Images

We also further tested a more challenging problem—super-resolving noisy images. The difficulty stems from the contradictory requirement that noise must be removed while the structural details of LR images, such as edges or textures, must be enhanced. Any inappropriate operations may enhance the noise or tend to reduce details. The objective evaluation results on super-resolving noisy images are shown in Table V. Our proposed method is more robust and still achieves consistently better performance than A+ and SRCNN in such challenging cases. As the noise level was increased from 3 to 7, the performance gap between the proposed method and two other methods widens from 1.41dB (A+) and 0.29dB (SRCNN) to 3.22dB (A+) and 1.14dB (SRCNN) in terms of PSNR.

G. Complexity Analysis

To evaluate the computational cost of the proposed method, we compared the running times of different methods on 12 images rescaled to 256 × 256 pixel on the 2 × image enlargement task. We calculated the average running time of our proposed method and five representative methods for these test images using MATLAB 2014a running on a computer with an Intel (R) Core (TM) i5-3230@2.60GHz and a 64-bit Windows 7 operating system. The global-stage content-based external image selection was implemented with Lire [62], which searches for similar images within 2 seconds. The local-stage high-frequency external patch matching was implemented based on the FLANN, which achieves fast patch retrieval and costs at most 34 seconds to search similar patches based on the joint vector. In all, the image search step introduces little additional computational burden. Group sparse coding does involve some computational penalty compared with previous sparse coding methods. The average computational time for an image enlargement (including the group sparse coding and non-local means) is illustrated in Table VI. As shown, our proposed approach improves SR performance at a cost of approximately 3 times the running cost of NCSR. The sparse representation methods (ScSR, ASDS, NCSR and the proposed method) are slower than A+ and SRCNN in the SR reconstruction phase because the framework of sparse representation is less efficient than anchor regression or CNN forward propagation. However, these two methods carry large burdens when training dictionaries, making it difficult for them to acquire external information adaptively from online training data for use in their dictionaries.

TABLE VI
THE AVERAGE RUNNING TIME OF DIFFERENT METHODS.

Method	**ScSR**	**ASDS**	**NCSR**
Dict. Training	hours	49.16s	68.20s
SR Reconstruction	183s	156.14s	278.18s
Method	**A+**	**SRCNN**	**Proposed**
Dict. Training	1281.76s	days	180.37s
SR Reconstruction	1.61s	12.33s	658.23s

There are some potential ways that our method can be accelerated. First, our method (and the other comparison methods for the image reconstruction) were all implemented with MATLAB. The speed of all these methods could be improved by implementing them in C++. In fact, our method could benefit further from pre-training some external sub-dictionaries and processing every image patch group in parallel. Second, some recent works [66, 67] have implemented sparse coding with a time-efficient learned feed-forward network. Implementing our proposed algorithm based on this learned framework might be a good choice to reduce the running time.

VII. Conclusions and Discussion

This paper presented a group structured sparse representation model that employs both external and internal similarities for image SR. Externally compensated correlated information is introduced by a two-stage retrieval and refinement process.

TABLE IV
PSNR (dB) results of the proposed method with references in different noise levels in 3× enlargement.

Noise Level	*Bike*	*Butterfly*	*Car*	*Comic*	*Field*	*Foreman*	*Hat*	*Leaves*	*Lena*	*Baboon*	*Text*	*Zebra*
0	25.01	28.83	29.65	24.86	25.57	32.10	31.60	28.28	32.31	21.75	14.54	25.81
20	24.99	28.81	29.58	24.85	25.56	32.16	31.58	28.31	32.36	21.75	14.54	25.83
50	24.99	28.74	29.58	24.84	25.61	32.17	31.56	28.28	32.32	21.75	14.53	25.83
100	24.99	28.82	29.64	24.84	25.62	32.18	31.58	28.28	32.32	21.75	14.54	25.82
150	24.99	28.81	29.62	24.84	25.59	32.16	31.59	28.28	32.33	21.75	14.54	25.81
200	24.98	28.82	29.62	24.84	25.60	32.16	31.55	28.28	32.37	21.75	14.56	25.81
300	24.94	28.80	29.54	24.84	25.59	32.05	31.51	28.24	32.27	21.75	14.53	25.81
Pure noise (100)	24.85	28.60	29.44	24.80	25.50	31.97	31.50	28.17	32.22	21.82	14.44	25.68
VAR(10^{-4})	26	60	47	3	14	55	14	18	23	6	13	24

TABLE V
PSNR (dB) and SSIM results when super-resolving noise images in 3× enlargement.

Noise Level	Method	A+		SRCNN		Proposed		Method	A+		SRCNN		Proposed	
	Image	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	Image	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3	*Bike*	23.56	0.7110	24.35	0.7619	24.37	0.7715	*Hat*	28.52	0.6792	30.07	0.7718	30.41	0.8143
5		22.72	0.6349	23.84	0.7177	24.10	0.7507		26.18	0.5239	28.71	0.6593	29.73	0.7827
7		21.59	0.5572	23.23	0.6579	23.73	0.7274		24.16	0.4123	27.20	0.5481	28.74	0.7456
3	*Butterfly*	25.45	0.8020	27.33	0.8587	27.55	0.8877	*Leaves*	24.57	0.8288	25.91	0.8764	26.99	0.9017
5		24.13	0.7047	26.48	0.7927	27.23	0.8748		23.40	0.7569	25.25	0.8306	26.64	0.8895
7		22.72	0.6268	23.23	0.6579	26.56	0.8621		22.22	0.6909	24.39	0.7822	26.00	0.8752
3	*Car*	27.20	0.7322	28.57	0.8202	28.88	0.8465	*Lena*	29.09	0.7211	30.87	0.8006	31.03	0.8289
5		25.35	0.5864	27.59	0.7247	28.54	0.8235		26.54	0.5751	27.36	0.7252	30.48	0.7965
7		23.58	0.4666	26.36	0.6248	28.10	0.7989		24.31	0.4576	27.64	0.6025	29.24	0.7668
3	*Comic*	23.43	0.7020	24.24	0.7533	24.36	0.7659	*Baboon*	21.28	0.4378	21.59	0.4829	21.59	0.4796
5		22.60	0.6287	23.80	0.7082	24.00	0.7421		20.71	0.3820	21.35	0.4508	21.42	0.4527
7		21.51	0.5546	23.15	0.6538	23.57	0.7172		20.08	0.3348	21.03	0.4172	21.27	0.4289
3	*Field*	24.37	0.5947	25.11	0.6798	25.16	0.7020	*Text*	13.15	0.6915	14.31	0.7807	13.76	0.6871
5		23.32	0.4746	24.64	0.5948	25.03	0.6814		13.10	0.6562	14.26	0.7572	14.01	0.6866
7		22.09	0.3782	23.98	0.5096	24.75	0.6599		13.01	0.6137	14.22	0.7296	13.90	0.6754
3	*Foreman*	28.55	0.7500	29.29	0.8404	31.03	0.8838	*Zebra*	24.28	0.6861	25.18	0.7420	25.14	0.7388
5		26.24	0.5924	28.14	0.7249	30.58	0.8589		23.24	0.6066	24.65	0.6977	25.01	0.7133
7		24.14	0.4648	26.85	0.6153	29.58	0.8424		21.99	0.5305	22.86	0.6421	24.61	0.6867
3								**Average**	24.45	0.6947	25.57	0.7641	**25.86**	**0.7757**
5									23.13	0.5935	24.67	0.6987	**25.56**	**0.7544**
7									21.78	0.5073	23.86	0.6257	**25.00**	**0.7322**

The content features in the global stage and the high frequency patch features in the local stage are jointly used to improve the selection process and refine the external information. The nonlocal redundancy is incorporated into the sparse representation model to form a group sparsity framework on an adaptively generated over-complete dictionary. This model is computationally highly efficient and thus convenient for absorbing external information dynamically. Based on the two-stage external information selection and the structured group sparse representation model, we exploit both the internal and retrieved external information to build an iterative integrated framework to super-resolve images. Experimental results demonstrate the superiority of our proposed method in using the complementary advantages of both the internal and external priors compared with state-of-the-art methods. It is interesting to observe that random noises may benefit learning a more expressive dictionary in some cases. This also motivates us to revisit the structured dictionary, including its advantages, drawbacks and potential capacities, in our future work.

References

[1] R. Y. Tsai and T. S. Huang, “Multipleframe image restoration and registration,” in *In Advances in Computer Vision and Image Processing*, vol. 1. JAI Press Inc.

[2] X. Li and M. T. Orchard, “New edge-directed interpolation,” *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521–7, 2001.

[3] L. Zhang and X. Wu, “An edge-guided image interpolation algorithm via directional filtering and data fusion,” *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2226–38, 2006.

[4] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, “Gradient profile prior and its applications in image super-resolution and enhancement,” *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1529–1542, June 2011.

[5] W. Zuo, L. Zhang, C. Song, and D. Zhang, “Texture enhanced image denoising via gradient histogram preservation,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, June 2013, pp. 1203–1210.

[6] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, “From local kernel to nonlocal multiple-model image denoising,” *Int’l Journal of Computer Vision*, vol. 86, no. 1, pp. 1–32, January 2010.

[7] W. Dong, L. Zhang, G. Shi, and X. Li, “Nonlocally centralized sparse representation for image restoration,” *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, April 2013.

[8] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, “Non-local sparse models for image restoration,” in *Proc. IEEE Int’l Conf. Computer Vision*, Sept 2009, pp. 2272–2279.

[9] A. Marquina and S. J. Osher, “Image super-resolution by TV-regularization and bregman iteration,” *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, December 2008.

[10] H. Aly and E. Dubois, “Image up-sampling using total-variation regularization with a new observation model,” *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1647–1659, Oct 2005.

[11] D. Glasner, S. Bagon, and M. Irani, “Super-resolution from a single image,” in *Proc. IEEE Int’l Conf. Computer Vision*, Sept 2009, pp. 349–356.

[12] G. Freedman and R. Fattal, “Image and video upscaling from local self-examples,” *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–10, 2010.

[13] C.-Y. Yang, J.-B. Huang, and M.-H. Yang, “Exploiting self-similarities for single frame super-resolution,” in *Proc. IEEE Asia Conf. Computer Vision*, 2011, pp. 497–510.

[14] Z. Zhu, F. Guo, H. Yu, and C. Chen, “Fast single image super-resolution via self-example learning and sparse representation,” *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2178–2190, Dec 2014.

[15] M.-C. Yang and Y.-C. F. Wang, “A self-learning approach to single image super-resolution,” *IEEE Transactions on Multimedia*, vol. 15, no. 3, pp. 498–508, 2013.

[16] J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in *IEEE Conference on Computer Vision and Pattern Recognition)*, 2015.

[17] H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, vol. 1, June 2004, pp. I–I.

[18] R. Timofte, V. De, and L. Van Gool, “Anchored neighborhood regression for fast example-based super-resolution,” in *Proc. IEEE Int’l Conf. Computer Vision*, Dec 2013, pp. 1920–1927.

[19] R. Timofte, V. De Smet, and L. Van Gool, “A+: Adjusted anchored

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2016.2614427, IEEE Transactions on Multimedia

neighborhood regression for fast super-resolution," *Proc. IEEE Asia Conf. Computer Vision*, pp. 111–126, 2014.

[20] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, June 2010.

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

[22] A. Chakrabarti, A. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel pca-based prior," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 888–892, June 2007.

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

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

[25] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467–3478, Aug 2012.

[26] 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. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2012, pp. 2216–2223.

[27] 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. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2013, pp. 345–352.

[28] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., 2014, vol. 8692, pp. 184–199.

[29] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. IEEE Int'l Conf. Computer Vision*, June 2015.

[30] Y. Zhu, Y. Zhang, and A. Yuille, "Single image super-resolution using deformable patches," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2014, pp. 2917–2924.

[31] L. Sun and J. Hays, "Super-resolution from internet-scale scene matching," in *IEEE International Conference on Computational Photography*, April 2012, pp. 1–12.

[32] J. Sun, J. Zhu, and M. Tappen, "Context-constrained hallucination for image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 231–238.

[33] R. Timofte, V. De Smet, and L. Van Gool, "Semantic super-resolution: When and where is it useful?" *Computer Vision and Image Understanding*, vol. 142, pp. 1–12, 2016.

[34] H. Yue, X. Sun, J. Yang, and F. Wu, "Landmark image super-resolution by retrieving web images," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4865–4878, Dec 2013.

[35] M. Zontak and M. Irani, "Internal statistics of a single natural image," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2011, pp. 977–984.

[36] I. Mosseri, M. Zontak, and M. Irani, "Combining the power of internal and external denoising," in *Proc. IEEE Int'l Conf. Computational Photography*, April 2013, pp. 1–9.

[37] H. Burger, C. Schuler, and S. Harmeling, "Learning how to combine internal and external denoising methods," in *Pattern Recognition*, 2013, vol. 8142, pp. 121–130.

[38] R. Timofte, R. Rothe, and L. Van Gool, "Seven ways to improve example-based single image super resolution," *arXiv preprint arXiv:1511.02228*, 2015.

[39] Z. Wang, Y. Yang, Z. Wang, S. Chang, J. Yang, and T. Huang, "Learning super-resolution jointly from external and internal examples," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4359–4371, Nov 2015.

[40] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[41] B. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 187–200, Jan 1999.

[42] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.

[43] J.-F. Cai, R. Chan, L. Shen, and Z. Shen, "Simultaneously inpainting in image and transformed domains," *Numerische Mathematik*, vol. 112, no. 4, pp. 509–533, 2009.

[44] J. Ren, J. Liu, and Z. Guo, "Context-aware sparse decomposition for image denoising and super-resolution," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1456–1469, April 2013.

[45] S. Bengio, F. Pereira, Y. Singer, and D. Strelow, "Group sparse coding," in *Proc. Annual Conference on Neural Information Processing Systems*, 2009, pp. 82–89.

[46] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[47] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation: Part i: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, March 2006.

[48] ——, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, 2006.

[49] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[50] J. Zhang, D. Zhao, F. Jiang, and W. Gao, "Structural group sparse representation for image compressive sensing recovery," in *Data Compression Conference*, March 2013, pp. 331–340.

[51] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, Aug 2014.

[52] J. Sun, Q. Chen, S. Yan, and L. F. Cheong, "Selective image super-resolution," *CoRR*, 2010.

[53] M. Eisemann, E. Eisemann, H.-P. Seidel, and M. Magnor, "Photo zoom: High resolution from unordered image collections," in *Proceedings of Graphics Interface*. Toronto, Ont., Canada, Canada: Canadian Information Processing Society, 2010, pp. 71–78.

[54] W. Bai, S. Yang, J. Liu, J. Ren, and Z. Guo, "Image super resolution using saliency-modulated context-aware sparse decomposition," in *Proc. IEEE Visual Communication and Image Processing*, 2013.

[55] C. Iakovidou, N. Anagnostopoulos, A. Kapoutsis, Y. Boutalis, and S. Chatzichristofis, "Searching images with MPEG-7 powered localized descriptors: The SIMPLE answer to effective content based image retrieval," in *12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, June 2014, pp. 1–6.

[56] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[57] S. A. Chatzichristofis and Y. S. Boutalis, "CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," in *Proceedings of the 6th International Conference on Computer Vision Systems*, 2008, pp. 312–322.

[58] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, vol. 2, June 2005, pp. 524–531 vol. 2.

[59] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, September 1975.

[60] M. Muja and D. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2227–2240, Nov 2014.

[61] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4544–4556, Nov 2012.

[62] SPAMS: a sparse modeling software. [Online]. Available: http://spams-devel.gforge.inria.fr/doc/html/index.html

[63] W. Dong, D. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, July 2011.

[64] H. Yue, X. Sun, J. Yang, and F. Wu, "Cloud-based image coding for mobile devices toward thousands to one compression," *IEEE Transactions on Multimedia*, vol. 15, no. 4, pp. 845–857, June 2013.

[65] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *ACM International Conference on Multimedia*. New York, NY, USA: ACM, 2014, pp. 675–678.

[66] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of International Conference on Machine Learning*, 2010, pp. 399–406.

[67] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *IEEE International*

Conference on Computer Vision, Dec 2015, pp. 370–378.

Jiaying Liu (S'09-M'10) received the B.E. degree in computer science from Northwestern Polytechnic University, Xian, China, and the Ph.D. degree with the Best Graduate Honor in computer science from Peking University, Beijing, China, in 2005 and 2010, respectively. She is currently an Associate Professor with the Institute of Computer Science and Technology, Peking University. She has authored or co-authored over 60 papers and 8 granted patents. Her current research interests include image processing, computer vision, and video compression. Dr. Liu was a Visiting Scholar with the University of Southern California, Los Angeles, from 2007 to 2008. She has visited Microsoft Research Asia (MSRA) since March 2015, supported by Star Track for Young Faculties. She has also served as TC member in APSIPA IVM since 2015, and APSIPA distinguished lecture in 2016-2017.

Wenhan Yang received the B.S degree in Computer Science from Peking University, Beijing, China, in 2008. He is currently a Ph.D. student with the Institute of Computer Science and Technology, Peking University. Mr. Yang was a Visiting Scholar with the National University of Singapore, from 2015 to 2016. His current research interests include image processing, sparse representation, image restoration and deep learning-based image processing.

Xinfeng Zhang (M'16) received the B.S. degree in computer science from Hebei University of Technology, Tianjin, China, in 2007, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014. He is currently a Research Fellow in Nanyang Technological University, Singapore. His research interests include image and video processing, image and video compression.

Zongming Guo (M'09) received the B.S. degree in mathematics, and the M.S. and Ph.D. degrees in computer science from Peking University, Beijing, China, in 1987, 1990, and 1994, respectively. He is currently a Professor with the Institute of Computer Science and Technology, Peking University. His current research interests include video coding, processing, and communication. Dr. Guo is the Executive Member of the China-Society of Motion Picture and Television Engineers. He was a recipient of the First Prize of the State Administration of Radio Film and Television Award in 2004, the First Prize of the Ministry of Education Science and Technology Progress Award in 2006, the Second Prize of the National Science and Technology Award in 2007, the Wang Xuan News Technology Award and the Chia Tai Teaching Award in 2008, the Government Allowance granted by the State Council in 2009, and the Distinguished Doctoral Dissertation Advisor Award of Peking University in 2012 and 2013.