Remote Sensing Image Super-Resolution Using Sparse Representation and Coupled Sparse Autoencoder

Zhenfeng Shao, Lei Wang, Zhongyuan Wang, and Juan Deng

***Abstract*—Remote sensing image super-resolution (SR) refers to a technique improving the spatial resolution, which in turn benefits to the subsequent image interpretation, e.g., target recognition, classification, and change detection. In popular sparse representation-based methods, due to the complex imaging conditions and unknown degradation process, the sparse coefficients of low-resolution (LR) observed images are hardly consistent with the real high-resolution (HR) counterparts, which leads to unsatisfactory SR results. To address this problem, a novel coupled sparse autoencoder (CSAE) is proposed in this paper to effectively learn the mapping relation between the LR and HR images. Specifically, the LR and HR images are first represented by a set of sparse coefficients, and then, a CSAE is established to learn the mapping relation between them. Since the proposed method leverages the feature representation ability of both sparse decomposition and CSAE, the mapping relation between the LR and HR images can be accurately obtained. Experimentally, the proposed method is compared with several state-of-the-art image SR methods on three real-world remote sensing image datasets with different spatial resolutions. The extensive experimental results demonstrate that the proposed method has gained solid improvements in terms of average peak signal-to-noise ratio and structural similarity measurement on all of the three datasets. Moreover, results also show that with larger upscaling factors, the proposed method achieves more prominent performance than the other competitive methods.**

***Index Terms*—Coupled sparse autoencoder (CSAE), image super-resolution (SR), remote sensing image, sparse representation.**

Manuscript received September 7, 2018; revised December 26, 2018 and April 26, 2019; accepted June 21, 2019. This work was supported in part by the National Key Technologies Research and Development Program under Grants 2016YFE0202300 and 2016YFB0502603, in part by the National Natural Science Foundation of China under Grants 61671332, 41771452, and 41771454, in part by the Fundamental Research Funds for the Central Universities under Grants 2042016kf0179 and 2042016kf1019, in part by the Special Task of Technical Innovation in Hubei Province under Grant 2017AAA123, and in part by the Natural Science Fund of Hubei Province under Grant 2018CFA007. *(Corresponding author: Lei Wang.)*

Z. Shao is with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079 China, and also with the Collaborative Innovation Center for Geospatial Technology, Wuhan 430079, China (e-mail: shaozhenfeng@whu.edu.cn).

L. Wang is with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: wlei@whu.edu.cn).

Z. Wang is with the National Engineering Research Center for Multimedia Software, Wuhan 430079, China (e-mail: wzy_hope@163.com).

J. Deng is with the School of Electronic Information, Wuhan University, Wuhan 430072, China (e-mail: jdeng@whu.edu.cn).

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/JSTARS.2019.2925456

I. Introduction

REMOTE sensing image with meter or submeter spatial resolution has become publicly available over the last decades. In spite of this recent development, its spatial resolution cannot meet the growing demand on image for productions and applications. This has been a serious impediment to the further development and application of remote sensing technology, whereas, it is considerably costlier and difficult to improve the resolution of sensors. However, image super-resolution (SR) technique provides a low-cost and effective way to alleviate this problem. Image SR can break through the limitation of sensor's instinct resolution and the influence of atmosphere [1], [2], and can produce an image with a better quality and a higher resolution, which provides the basis for further image analyses and applications [3].

Over the past few decades, a series of image SR methods have been proposed, which can be roughly classified into three categories [4]: 1) interpolation-based approaches; 2) reconstruction-based approaches; and 3) learning-based approaches. Interpolation-based approaches are the most intuitive ones, which acquire the high-resolution (HR) images with linear, bilinear, or bicubic interpolation. However, the interpolation operation always blurs the image inevitably.

Reconstruction-based approaches consist of iterative back projection [5], projection on convex set (POCS) [6], maximum *a posteriori* (MAP), and regularization methods. Zhang *et al.* [7] used POCS to reconstruct the three-view remote sensing image. Luo *et al.* [8] modified the MAP and the variable Bayesian method to recover the remote sensing image and obtained considerably satisfactory results, although it involves large computation costs. Regularization methods [9] convert the original ill-posed problem to an optimization problem by introducing a total variation regularization term. The advantage of reconstruction-based approaches [10] is that some local *a priori* hypotheses are enough to relieve the blurring phenomena caused by interpolation. However, when upscaling factor is relatively larger, it is not easy to obtain accurate motion information of subpixels, which has a significant impact on the reconstruction results. Therefore, the hybrid method that combines the advantages of different reconstruction methods has also been widely used. However, due to the poor ability of reconstructing the subpixel motion information of the reconstruction methods, the hybrid method fails to satisfy the reconstruction accuracy.

Learning-based image SR approaches aim at mining the relationship between the low-resolution (LR) and the corresponding HR image patches with machine learning methods. They mainly include Markov network-based methods, neighborhood embedding methods, deep learning-based methods, and sparse representation-based methods. Markov network-based methods [11], [12] build a probability field for the HR image based on the given LR image, and then the HR image is reconstructed by maximizing the condition probability. Neighborhood embedding methods originate from the manifold learning, which assumes that there exist some similar geometry structures in the feature spaces of the LR and HR images. The corresponding HR image is, thus, predicted by constraining the smoothness between the central patch and its neighborhoods [13]. Deep learning-based methods [14]–[19] extract the image features with deep hidden layers of the network under the supervision of HR training samples. Benefiting from the deep network structure, deep learning methods can capture informative features of images for SR. However, to efficiently capture the desired image features, they always require a large amount of training samples and computing resources.

Actually, image signal can be represented as a linear combination of a series of basic structural elements from a redundant dictionary in a local range [20], [21]. Based on this, sparse representation [22] aims at decomposing the LR and HR image patches into sparse coefficients over the corresponding redundant dictionaries. By using the joint dictionary training strategy [20], the sparse coefficients of the LR and HR images are constrained to be consistent. Then, the LR sparse coefficients can be used to reconstruct the corresponding HR images according to the trained HR dictionary. Due to their promising performance, a lot of modified variants [23]–[26] have been proposed to suit various tasks. He *et al.* [27] trained the joint dictionary with beta process, which could remove some useless elements in the redundant dictionary. Instead of constraining the sparse coefficients of the LR and HR images to be equal, they established a weight matrix to map the LR coefficients to the HR coefficients. Peleg and Elad [28] utilized a statistical model to constrain the mapping between the LR and HR coefficients, and the dictionaries were then solved by minimizing an optimal problem. In addition, a cascade strategy is also used to improve the reconstruction precision. Yeganli *et al.* [29] and Zhang *et al.* [30] trained more distinctive targeted dictionaries, such as edges, gradients, and structures, so that the corresponding dictionary can be selected to produce better reconstruction results for specific kind of images.

Generally, most of the existing sparse representation-based image SR methods [31]–[34] used the joint dictionary training strategy or its variants to train the dictionaries, whose LR coefficients can be used as the HR coefficients directly or a weight matrix can be constructed to bridge them. However, due to the complex imaging conditions and unknown degradation process, the information contained in remote sensing images at different spatial scales (or resolutions) usually varies much (even for the same scene). As a result, the images at different spatial scales have different optimized sparse representations. The training strategy using simple linear mapping is difficult to reflect their complex corresponding relation, which has become an important limitation for further image SR.

In this paper, we propose a novel coupled sparse autoencoder (CSAE) to excavate the mapping relation between the LR and HR sparse coefficients using learning strategy. Meanwhile, since CSAE accepts the sparse coefficients as *a priori* knowledge to guide the learning process, the proposed method leverages the feature representation ability of both sparse decomposition and CSAE. Therefore, the HR coefficients can be accurately estimated from the given LR sparse coefficients and used for HR image reconstruction. Experimental results on three remote sensing image datasets demonstrate that the proposed method has achieved impressive reconstruction results at different spatial resolutions. In addition, compared to the joint dictionary training methods, the HR coefficients predicted by the proposed method also exhibit larger correlation with the true values.

The remainder of this paper is organized as follows. Section II introduces the preliminaries and problem setup for image SR. Section III describes the details of the proposed method. Experimental results and further discussions are then presented in Section IV. Finally, a conclusion of our work and a layout of future work are drawn in Section V.

II. Preliminaries and Problem Setup

The observed image can be regarded as an LR image obtained from the HR image through a series of degradation processes, including optical dimming, subsampling, and additive noise. Our goal is to reconstruct the corresponding HR image based on the observed LR image. Let I_h and I_l be the HR image and the LR image, respectively. The degradation model can be typically formulated as follows:

$$I_l = SBI_h + \tilde{n} \tag{1}$$

where S indicates the downsampling matrix, B represents the blurring matrix, and $\tilde{n}$ is the additive noise.

To reconstruct the HR image from the observed LR image, the key is to build the mapping relation between them. However, this is a much difficult task in practice due to the unknown degradation factors, whereas both the LR and HR images can be linearly represented by sparse coefficients over their corresponding redundant dictionaries in a local range. Thus, the relationship can be alternatively built between the sparse coefficients of the LR and HR images.

For a given LR image patch p_l (which is extracted from the LR image and has been flattened to a vector), it can be represented as the product of a redundant dictionary and the corresponding sparse coefficients

$$p_l \approx D_l \alpha_l \tag{2}$$

where D_l is the redundant dictionary of the LR image and α_l is the corresponding sparse coefficient of the LR image (SCOLRI) with most of its elements being zero. Similarly, the HR image patch p_h can be represented as $p_h \approx D_h \alpha_h$ as well. The sparse coefficients α_l and α_h can be regarded as the features of p_l and p_h, respectively, and the relation between α_l and α_h implies the relation between the LR and HR images.

The conventional method for dictionary training is the joint dictionary training strategy [20], which trains the LR and HR dictionaries together by solving the following optimization problem:

$$\min_{D,\alpha} \|p - D\alpha\|_2^2 + \lambda\|\alpha\|_0 \tag{3}$$

where $D = \begin{bmatrix} D_h \\ D_l \end{bmatrix}$ and $p = \begin{bmatrix} p_h \\ p_l \end{bmatrix}$. $\|\cdot\|_0$ indicates the l_0 norm and λ is the penalty factor of sparsity constraint. Joint dictionary training strategy trains the dictionaries by stacking the LR and HR image patches together so that the LR patch p_l and the HR patch p_h are constrained to share the same sparse coefficients over their corresponding dictionaries. During the process of image SR, the given LR image patch p_l is decomposed over the LR dictionary D_l to obtain the sparse coefficients α. The HR image patch is then predicted by a linear combination of the sparse coefficients α with the HR dictionary D_h, *i.e.*, $p_h = D_h\alpha$. Finally, the HR image is reconstructed by splicing the predicted HR image patches together.

However, as the LR and HR images have different frequency ranges, they actually need different number of basic structural elements for sparse representation. Joint dictionary training strategy constrains the sparse coefficients of the LR and HR images to have the same length, which makes the dictionaries of both the LR and HR images hardly represent their image spaces optimally. To address this problem, He *et al.* [27] trained the dictionaries with beta process, which utilized a weight matrix to reflect the mapping between the LR and HR coefficients. Nevertheless, no matter forcing the sparse coefficients of the LR and HR images to be equal or assuming there is a linear mapping relation, it is inadequate to characterize the real complex image SR process. Therefore, this paper aims at establishing a novel CSAE to accurately learn the mapping between the LR and HR sparse coefficients.

Notably, according to (1), we have the following equation:

$$BI_h = S^{-1}I_l - S^{-1}\tilde{n} \tag{4}$$

where S^{-1} is the inverse (or pseudoinverse) of matrix S, named interpolation matrix. $\tilde{I}_l = BI_h$ indicates the LR image, which is blurred from the HR image I_h or up-sampled from the LR image I_l. Denote the residual $I_{hl} = I_h - \tilde{I}_l = I_h - BI_n = (I - B)I_h$, where I is the identical matrix. Let $H = I - B$, we have $I_{hl} = HI_h$, and H is a high-pass matrix. Therefore, the residuals (the differences between the HR image and the up-sampled LR image) contain the high-frequency components of the original HR image, such as textures, edges, and so on (see Fig. 1).

Accordingly, since the HR image has the same low-frequency information as the LR image, we only have to estimate their residual I_{hl}. The HR image can be then reconstructed by adding the estimated residuals to the up-sampled LR image, *i.e.*, $I_h = I_{hl} + \tilde{I}_l$.

III. SR With Sparse Decomposition and CSAE

In this paper, a novel CSAE is proposed to learn the mapping relation between the sparse coefficients of the LR image and the residuals. When the training is done, the sparse coefficients of the residuals (SCOR) can be predicted by the proposed CSAE, and then the residuals can be further estimated. At last, the final HR image is reconstructed by adding the estimated residuals to the up-sampled LR image (see Fig. 2).

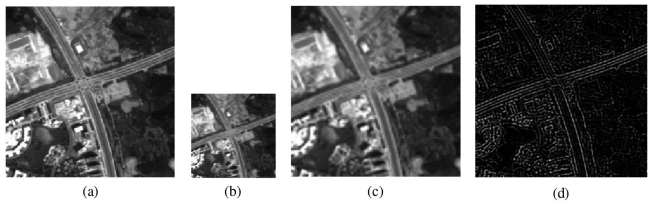


Fig. 1. Information contained in the residuals. (a) Original HR image. (b) LR image. (c) Up-sampled LR image. (d) Visual display of their residuals (difference between the HR image and the up-sampled LR image). The residuals contain both positive and negative pixel values, we show it with scaling for visual convenience.

A. Image Sparse Decomposition

This section will describe how to decompose the patches of the LR images and the residuals to sparse coefficients. Denote the LR image patches as $P_l = [p_l^{(1)}, p_l^{(2)}, \ldots]$, the corresponding redundant dictionary of LR image D_l can be trained with the following optimization problem:

$$\min_{D_l,\alpha_l} \|P_l - D_l\alpha_l\|_2^2 + \lambda_s\|\alpha_l\|_0 \tag{5}$$

where α_l indicates the corresponding SCOLRI. It is an NP-hard problem with l_0 norm constraint, and a greedy algorithm is needed to approach the approximate solution. Nevertheless, Donoho [35] has shown that when the dictionary is redundant enough, this l_0 norm problem is equivalent to the following l_1 norm problem:

$$\min_{D_l,\alpha_l} \|P_l - D_l\alpha_l\|_2^2 + \lambda_s\|\alpha_l\|_1 \tag{6}$$

where λ_s is the penalty factor of sparsity constraint. Dictionary D_l can be trained by rich image samples with K-SVD [36] algorithms and the corresponding sparse coefficient α_l can be solved simultaneously.

Similarly, the patches of the residuals $P_{hl} = [p_{hl}^{(1)}, p_{hl}^{(2)}, \ldots]$ can also be decomposed with the optimization problem as follows:

$$\min_{D_{hl},\alpha_{hl}} \|P_{hl} - D_{hl}\alpha_{hl}\|_2^2 + \lambda_s\|\alpha_{hl}\|_1 \tag{7}$$

where α_{hl} is the sparse coefficient of the residuals and D_{hl} is the corresponding dictionary. α_{hl} and D_{hl} are solved by minimizing (7) with K-SVD algorithms.

After that, the decomposed sparse coefficients of the LR image and the residuals (i.e., α_l and α_{hl}) are used as the training samples for the proposed CSAE (see Section III-B).

B. Mapping Learning With CSAE

Image SR is to restore the high-frequency components of the LR image, which equals to estimate SCOR $\alpha_{hl} \in R^{d_{hl}}$ from the given SCOLRI $\alpha_l \in R^{d_l}$. To this end, many related methods

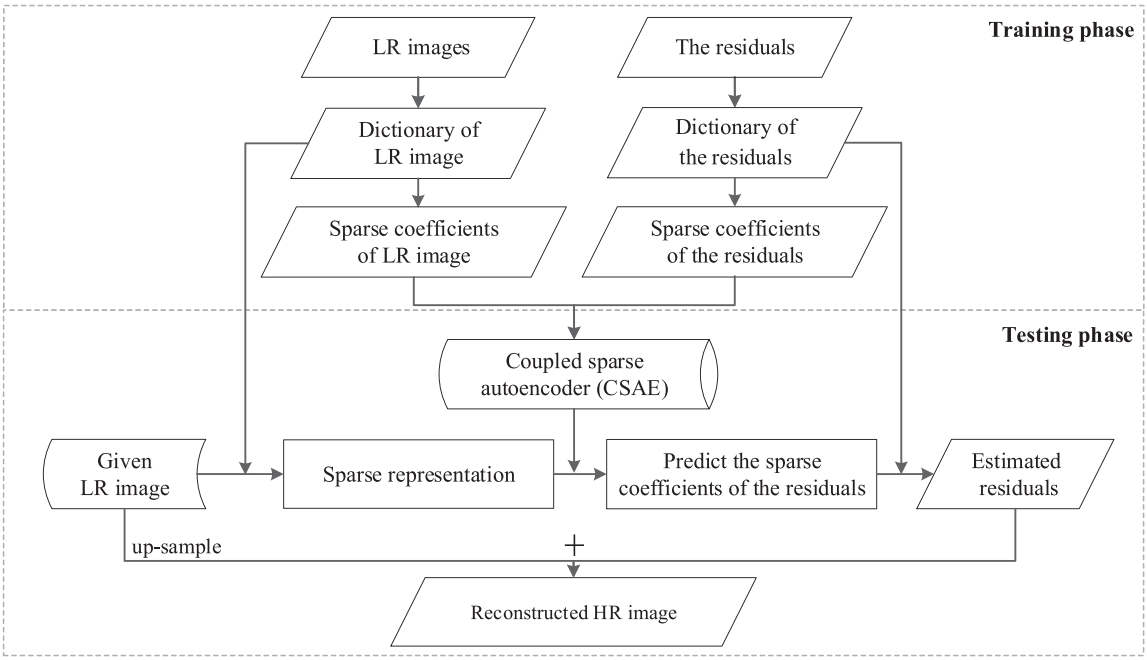


Fig. 2. Framework of the proposed method for image super-resolution. During the training phase, the LR images and the residuals (the differences between the HR and up-sampled LR images) are represented as sparse coefficients, and the CSAE is trained to learn their mapping relation. During the testing phase, we first represent the given LR image as sparse coefficients, and the sparse coefficients of the residuals can be predicted using the learned CSAE. The residuals are then estimated according to their predicted sparse coefficients, and the final HR image is reconstructed by adding the estimated residuals to the up-sampled LR image.

have been proposed [27]–[30]. However, most of them are linear mapping, which cannot accurately reflect the relation between the low and high frequencies. In addition, due to the discrete sampling and noise, the linear mapping is difficult to characterize the internal mechanism of image SR.

To address these problems, a CSAE aiming at learning the mapping relation between α_l and α_{hl} is proposed in this paper. Specifically, CSAE maps α_l and α_{hl} to a hidden feature space, where they are constrained to be equal. Therefore, we find a transition feature space, where the hidden representations of α_l and α_{hl} are equal to each other. Map the SCOLRI to its hidden representation in the feature space and the SCOR can be then reconstructed from it. In addition, to guarantee the sparsity of the reconstructed result, sparsity constraint is needed during each phase of the mapping.

For the given SCOLRI $\alpha_l \in R^{d_l}$, its hidden feature representation is obtained by using a sparse autoencoder (SAE). SAE is a kind of neural network consisting of a hidden layer and constraints the output to be equal to the input. It consists of two parts: encoder and decoder. The encoder maps the input α_l to a hidden feature representation $\alpha_l^{(h)} \in R^h$, while the decoder reconstructs α_l from $\alpha_l^{(h)}$. Denote the reconstructed α_l as $\hat{\alpha}_l$, $\hat{\alpha}_l$ is expected to be equal with α_l.

Technically, denote $\mathrm{Encoder}_l$ and $\mathrm{Decoder}_l$ as the encode and decode function for α_l. The hidden feature representation of α_l can be calculated as

$$\alpha_l^{(h)} = \mathrm{Encoder}_l\left(\alpha_l\right) \tag{8}$$

and the reconstructed α_l can be formulated as

$$\hat{\alpha}_l = \mathrm{Decoder}_l\left(\alpha_l^{(h)}\right) = \mathrm{Decoder}_l\left(\mathrm{Encoder}_l\left(\alpha_l\right)\right). \tag{9}$$

The loss function to train the SAE (including the encoder and the decoder) is

$$L_l = \frac{1}{2}\left\|\alpha_l - \hat{\alpha}_l\right\|_2^2 + \lambda_l\left\|\alpha_l^{(h)}\right\|_1 \tag{10}$$

where λ_l indicates the factor of sparsity constraint to keep the sparsity of output results. On one hand, we constraint α_l and $\hat{\alpha}_l$ close to each other. On the other hand, the hidden feature $\alpha_l^{(h)}$ is also expected to remain the sparsity as α_l.

Similarly, we can also establish the corresponding $\mathrm{Encoder}_{hl}$ and $\mathrm{Decoder}_{hl}$ for α_{hl} which satisfies

$$\alpha_{hl}^{(h)} = \mathrm{Encoder}_{hl}\left(\alpha_{hl}\right) \tag{11}$$

$$\hat{\alpha}_{hl} = \mathrm{Decoder}_{hl}\left(\alpha_{hl}^{(h)}\right) \tag{12}$$

where $\alpha_{hl}^{(h)} \in R^h$ is the hidden feature of α_{hl}, $\hat{\alpha}_{hl}$ indicates the reconstruction of α_{hl}. The corresponding loss function can be formulated as

$$L_{hl} = \frac{1}{2}\left\|\alpha_{hl} - \hat{\alpha}_{hl}\right\|_2^2 + \lambda_{hl}\left\|\alpha_{hl}^{(h)}\right\|_1 \tag{13}$$

where λ_{hl} indicates the factor of sparsity constraint.

The hidden representation $\alpha_l^{(h)}$ and $\alpha_{hl}^{(h)}$ can be seen as the feature vector of α_l and α_{hl}, respectively. The proposed CSAE aims at coupling the SAEs of α_l and α_{hl} together so that the

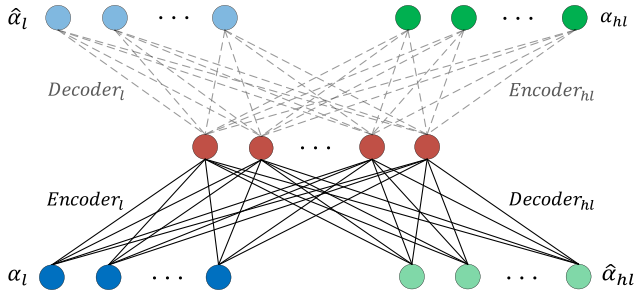


Fig. 3. Structure of the proposed CSAE. CSAE aims at learning the mapping relation between α_l and α_{hl}. Once the training of CSAE is done, $\mathrm{Decoder}_l$ and $\mathrm{Encoder}_{hl}$ (the dotted lines) are not needed. By first mapping the SCOLRI α_l to a hidden feature space with $\mathrm{Encoder}_l$, the corresponding SCOR can be estimated from the hidden feature representation with $\mathrm{Decoder}_{hl}$.

mapping relation between α_l and α_{hl} can be established. Specifically, CSAE constrains their hidden representations in the feature space to be equal, *i.e.*, $\alpha_l^{(h)} = \alpha_{hl}^{(h)}$. Therefore, the loss function of the proposed CSAE can be represented as follows:

$$\mathrm{Loss} = L_l + L_{hl} + L_c \tag{14}$$

where $L_c = \frac{1}{2}\|\alpha_l^{(h)} - \alpha_{hl}^{(h)}\|_2^2$ indicates the coupling loss, which constrains the hidden representations $\alpha_l^{(h)}$ and $\alpha_{hl}^{(h)}$ to be equal. The parameters of CSAE are optimized by minimizing the loss function (14) using batch stochastic gradient descent with momentum.

Once the model is trained, we have $\alpha_l^{(h)} \approx \alpha_{hl}^{(h)}$, and the decode function of α_l and the encode function of α_{hl} are not needed (see Fig. 3). The mapping relation between α_l and α_{hl} can then be established with the trained $\mathrm{Encoder}_l$ and $\mathrm{Decoder}_{hl}$. For the given SCOLRI α_l, the corresponding estimated SCOR can be formulated as

$$\hat{\alpha}_{hl} = \mathrm{Decoder}_{hl}\left(\mathrm{Encoder}_l\left(\alpha_l\right)\right). \tag{15}$$

The output $\hat{\alpha}_{hl}$ indicates the reconstruction of α_{hl}. Equation (15) describes the reconstruction process of SCOR α_{hl} from SCOLRI α_l, which reflects the mapping relation between the LR and HR images.

Notably, the encoder and decoder of α_l and α_{hl} can be implemented with any differentiable architecture. In this paper, a neural network with one hidden layer is used as it already has the ability to appropriate any nonlinear function.

C. HR Image Reconstruction

This section describes how to reconstruct the HR image from the given LR image using sparse decomposition and the proposed CSAE (see Algorithm 1). For the given LR image I_l, we first slice it into overlapping patches $\mathcal{P}_l = \{p_l^{(1)}, p_l^{(2)}, \ldots\}$. For each LR image patch $p_l^{(k)} \in \mathcal{P}_l$, its sparse coefficients $\alpha_l^{(k)}$ are calculated by solving the following optimization problem:

$$\min_{\alpha_l^{(k)}} \left\| p_l^{(k)} - D_l\alpha_l^{(k)} \right\|_2^2 + \lambda_s \left\| \alpha_l^{(k)} \right\|_1 \tag{16}$$

Algorithm 1: Image SR With The Proposed CSAE.

Input: LR image I_l, dictionary D_l and D_{hl}, $\mathrm{Encoder}_l$ and $\mathrm{Decoder}_{hl}$, up-sample matrix S^{-1}
Output: Reconstructed HR image I_h
Slice the LR image I_l into overlapping patches $\mathcal{P}_l = \{p_l^{(1)}, p_l^{(2)}, \ldots\}$;
For each $p_l^{(k)} \in \mathcal{P}_l$ **do**
  Calculate the sparse coefficient $\alpha_l^{(k)}$ by solving (16);
  Estimate the corresponding sparse coefficients of the residuals $\alpha_{hl}^{(k)} = \mathrm{Decoder}_{hl}(\mathrm{Encoder}_l(\alpha_l^{(k)}))$;
  Calculate the corresponding residual patch $p_{hl}^{(k)} = D_{hl}\alpha_{hl}^{(k)}$;
  Reconstruct the corresponding HR image patch $p_h^{(k)} = p_{hl}^{(k)} + S^{-1}p_l^{(k)}$;
end for
Reconstruct the HR image $\hat{I}_h$ by splicing $\{p_{hl}^{(1)}, p_{hl}^{(2)}, \ldots\}$ together and averaging their overlaps;
Optimize the reconstruct HR image $\hat{I}_h$ to obtain the final HR image I_h by solving (19).

where D_l is the dictionary of LR image solved by minimizing (6). Then, the corresponding SCOR is estimated using the proposed CSAE as

$$\alpha_{hl}^{(k)} = \mathrm{Decoder}_{hl}\left(\mathrm{Encoder}_l\left(\alpha_l^{(k)}\right)\right). \tag{17}$$

The corresponding residuals are obtained as

$$p_{hl}^{(k)} = D_{hl}\alpha_{hl}^{(k)} \tag{18}$$

where D_{hl} is the dictionary of the residuals solved by minimizing (7). After that, the corresponding HR image patch $p_h^{(k)}$ is reconstructed by adding the estimated residuals to the up-sampled LR image, *i.e.*, $p_h^{(k)} = p_{hl}^{(k)} + S^{-1}p_l^{(k)}$. At last, the HR image $\hat{I}_h$ is obtained by splicing the reconstructed HR image patches together and averaging their overlaps.

Notably, due to the inference of noise, the reconstructed HR image $\hat{I}_h$ may not satisfy the constraint conditions completely. To remove the vagueness, the final HR image I_h is optimized by the following global optimization problem:

$$I_h = \mathrm{argmin}_{I_h} \|SBI_h - I_l\|_2^2 + \mu \left\| I_h - \hat{I}_h \right\|_2^2 \tag{19}$$

where μ is the factor of the regularization term. It shows that the final HR image after blurring and subsampling should be close to the observed LR image as possible.

IV. Experimental Results and Discussion

A. Experimental Datasets and Settings

To verify the validity of the proposed method, three groups of remote sensing images with different spatial resolutions are applied in our experiments. They are NWPU VHR-10 [37] image set with 1-m spatial resolution, ZY-3 images with 5.6-m spatial resolution, and MOMs-2P images with 18-m spatial resolution. Each image set is split into the training set and the testing set,

accounting for 90% and 10%, respectively. For each image set, the testing images are randomly selected from the testing set. Conveniently, the five randomly selected NWPU VHR-10 [37] images are denoted as Im1, Im2, …, Im5, the five ZY-3 images are denoted as Im6, Im7, …, Im10, and the five MOM-2P images are Im11, Im12, …, Im15. The experimental results are quantitatively evaluated with two commonly used indexes: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

For the convenience of accuracy assessment, the original image I_h of each image set is regarded as the HR image. The LR image $\tilde{I}_l$ is generated from I_h by the downsampling operator, and then interpolated back to have the same size as I_h with the given scale factor using bicubic algorithm. The residual I_{hl} is calculated as $I_{hl} = I_h - \tilde{I}_l$. The first- and second-order derivatives are utilized as the training features instead of the LR image itself. According to Yang *et al.* [20], the gradient operators

$$\begin{cases} f_1 = [-1, 0, 1], & f_2 = f_1^T \\ f_3 = [1, 0, -2, 0, 1], & f_4 = f_3^T \end{cases} \tag{20}$$

are used in this paper to get the first- and second-order derivatives of the LR images. By applying these four gradient operators to the LR image respectively, we obtain four derivatives of each image and combine them together as training samples.

After that, 100 000 image patches with size 7×7 pixels are randomly chosen from the training set for dictionary training. The dictionaries are trained using K-SVD algorithm with $\lambda_s = 0.15$ (see Section III-A). As the residuals contain only a small part of information of the HR image, the dictionary sizes of the LR and the residuals are set to 256 and 64, respectively. Computing of the residuals also reduces the prediction space and the difficulties. Notably, the proposed method aims at reconstructing the HR image from a single image with one channel. For RGB color image, we first converted it into YUV color space, and only reconstructed the HR image at Y channel. The final HR image is then obtained by converting the results back to RGB color space. In addition, to further explore the effects of different upscaling factors, we conducted two group of experiments with scale factor $s = 2$ and $s = 3$ while maintaining the other experimental settings unchanged.

During the training phase of CSAE, the neuron number of the hidden layer is set to 192 and the sparsity constraint factor is $\lambda_l = \lambda_{hl} = 0.1$. The loss function in (14) is optimized using stochastic batch gradient descent algorithm with batch size 100. In addition, to deeply understand our CSAE, the performance of several key components is also further discussed and analyzed individually.

B. Comparisons With Other Methods

We compared our SR results with other state-of-the-art image SR methods, including compressive sensing with a redundant dictionary (CSRD) [26], beta process joint dictionary learning (BPJDL) [27], sparse structural manifold embedding (SSME) [24], and FSRCNN [38].

1) Experiments on the NWPU VHR-10 Image Set: The experimental results on the NWPU VHR-10 image set with scale factor $s = 2$ and $s = 3$ are provided in Tables I and II, respectively. The results show that the proposed method has achieved the highest PSNR and SSIM measures on most of the testing images. In addition, the larger the upscaling factor is, the more high-frequency information is missed, which results in that it is more difficult to construct the mapping relation between the LR and HR images.

TABLE I
PSNR (dB) AND SSIM RESULTS USING DIFFERENT METHODS ON THE NWPU VHR-10 IMAGE SET WITH SCALE FACTOR $s = 2$

	Index	CSRD [26]	BPJDL [27]	SSME [24]	FSRCNN [38]	Proposed
Im1	PSNR	32.35	32.86	31.78	32.73	**33.04**
	SSIM	0.973	0.977	0.944	**0.982**	**0.982**
Im2	PSNR	22.31	22.67	21.11	**22.87**	22.64
	SSIM	0.736	0.772	0.736	**0.790**	0.779
Im3	PSNR	25.67	26.09	25.03	**26.19**	26.16
	SSIM	0.895	0.910	0.832	0.938	**0.958**
Im4	PSNR	29.56	30.23	29.54	30.24	**30.29**
	SSIM	0.966	**0.977**	0.920	0.975	0.976
Im5	PSNR	32.56	33.17	32.60	33.02	**33.22**
	SSIM	0.971	0.975	0.921	0.975	**0.977**
Average	PSNR	28.488	29.006	28.014	29.030	**29.070**
	SSIM	0.9085	0.9224	0.8705	0.9320	**0.9343**

TABLE II
PSNR (dB) AND SSIM RESULTS USING DIFFERENT METHODS ON THE NWPU VHR-10 IMAGE SET WITH SCALE FACTOR $s = 3$

	Index	CSRD [26]	BPJDL [27]	SSME [24]	FSRCNN [38]	Proposed
Im1	PSNR	29.12	29.24	26.88	29.31	**29.43**
	SSIM	0.960	0.961	0.900	0.948	**0.963**
Im2	PSNR	20.06	20.16	19.07	20.17	**20.22**
	SSIM	0.604	0.610	0.581	**0.630**	0.613
Im3	PSNR	22.92	23.07	22.39	23.16	**23.17**
	SSIM	0.715	0.718	0.710	0.711	**0.718**
Im4	PSNR	26.14	26.33	26.05	**26.64**	26.54
	SSIM	0.939	0.941	0.853	0.934	**0.943**
Im5	PSNR	29.66	29.80	28.75	29.71	**29.88**
	SSIM	0.835	0.838	0.833	0.835	**0.840**
Average	PSNR	25.580	25.721	24.628	25.800	**25.850**
	SSIM	0.8106	0.8136	0.7753	0.8110	**0.8155**

To evaluate the reconstruction results intuitively, Figs. 4 and 5 show some reconstructed NWPU VHR-10 images with the proposed and other competitive image SR methods. It can be seen that the proposed method can well maintain the image structures and details. By combining the sparse representation with CSAE, the proposed method can excavate the mapping relation between the sparse coefficients of the LR image and the residuals more accurately, keeping the structure and texture information well.

2) Experiments on the ZY-3 Image Set: Tables III and IV show the experimental results on the ZY-3 images with scale factor $s = 2$ and $s = 3$, respectively. In general, our performance is on par with or better than other competitive methods on most of the testing images. In addition, compared to the results on the

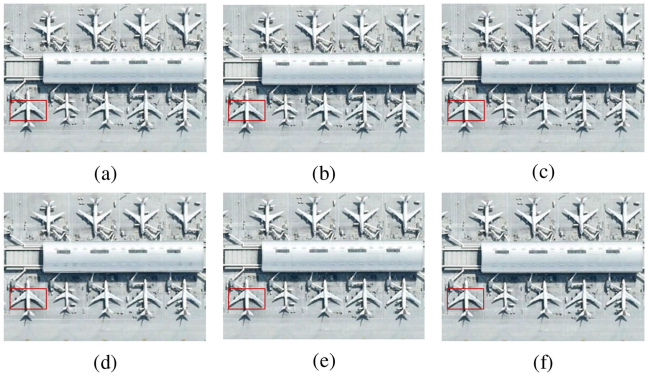


Fig. 4. Reconstructed NWPU VHR-10 images with scale factor $s = 2$. (a) Original image. (b) CSRD [26]. (c) BPJDL [27]. (d) SSME [24]. (e) FSRCNN [38]. (f) Proposed method.

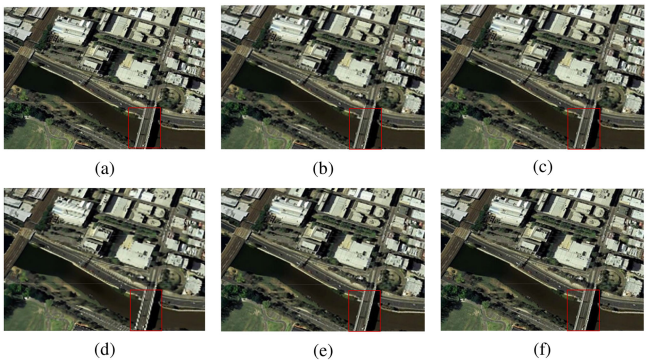


Fig. 5. Reconstructed NWPU VHR-10 images with scale factor $s = 3$. (a) Original image. (b) CSRD [26]. (c) BPJDL [27]. (d) SSME [24]. (e) FSRCNN [38]. (f) Proposed method.

TABLE III
PSNR (dB) AND SSIM RESULTS USING DIFFERENT METHODS ON ZY-3 IMAGES WITH SCALE FACTOR $s = 2$

	Index	CSRD [26]	BPJDL [27]	SSME [24]	FSRCNN [38]	Proposed
Im6	PSNR	37.84	38.81	38.26	38.85	**39.18**
	SSIM	0.977	0.982	0.975	0.980	**0.983**
Im7	PSNR	35.77	36.68	36.82	**37.19**	37.00
	SSIM	0.973	**0.980**	0.975	0.979	**0.980**
Im8	PSNR	34.40	35.68	35.66	36.01	**35.73**
	SSIM	0.962	**0.971**	0.970	0.972	**0.971**
Im9	PSNR	34.59	35.05	34.67	35.43	**36.14**
	SSIM	0.969	0.974	0.970	0.974	**0.977**
Im10	PSNR	38.54	39.01	39.12	39.69	**40.46**
	SSIM	0.982	0.986	0.980	0.985	**0.988**
Average	PSNR	36.228	37.046	36.906	37.435	**37.702**
	SSIM	0.9726	0.9786	0.974	0.9780	**0.9797**

NWPU VHR-10 image set, the ZY-3 images have lower spatial resolution, and the relation between the LR and the original HR image is relatively simple and thus leads to better reconstruction results. Compared to the NWPU VHR-10 images, the average PSNR of the proposed method with MOM-2P images

TABLE IV
PSNR (dB) AND SSIM RESULTS USING DIFFERENT METHODS ON ZY-3 IMAGES WITH SCALE FACTOR $s = 3$

	Index	CSRD [26]	BPJDL [27]	SSME [24]	FSRCNN [38]	Proposed
Im6	PSNR	32.52	32.92	32.44	33.02	**33.35**
	SSIM	0.922	0.930	0.913	0.928	**0.934**
Im7	PSNR	29.76	29.40	30.07	**30.76**	30.12
	SSIM	0.898	0.900	0.903	**0.913**	**0.907**
Im8	PSNR	29.56	29.66	29.06	30.18	**30.26**
	SSIM	0.881	0.886	0.879	0.894	**0.898**
Im9	PSNR	29.76	30.13	29.45	**30.33**	30.32
	SSIM	0.910	0.915	0.913	0.917	**0.921**
Im10	PSNR	32.31	31.78	32.80	32.81	**33.40**
	SSIM	0.927	0.925	0.915	0.930	**0.941**
Average	PSNR	30.782	30.778	30.764	31.423	**31.489**
	SSIM	0.9076	0.9112	0.9046	0. 9165	**0.9201**

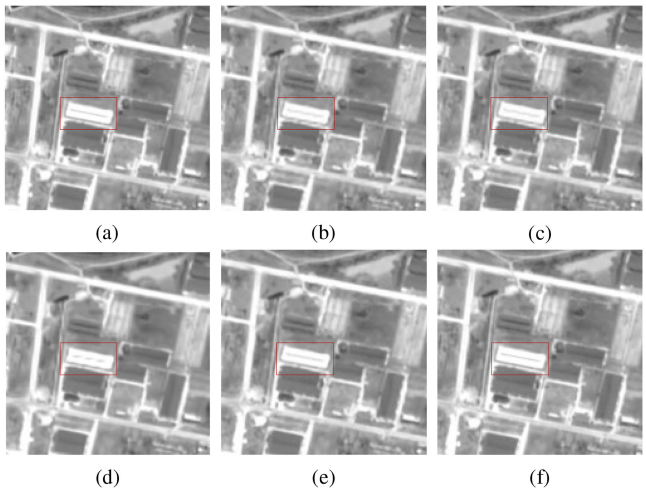


Fig. 6. Reconstructed ZY-3 images with scale factor $s = 2$. (a) Original image. (b) CSRD [26]. (c) BPJDL [27]. (d) SSME [24]. (e) FSRCNN [38]. (f) Proposed method.

gains 8.632 dB and 5.639 dB for scale factor $s = 2$ and $s = 3$, respectively.

Figs. 6 and 7 provide some reconstructed ZY-3 images with the proposed method and other competitive image SR methods. Similar as the experimental results on the NWPU VHR-10 image set, the proposed method has clearly reconstructed the structure and texture details of the images.

3) Experiments on the MOMs-2P Image Set: The experimental results on the MOM-2P images are provided in Tables V and VI. Results in Table V demonstrate that the proposed method has achieved the highest PSNR and SSIM measures on average, and Table VI provides the similar results as Table V but even better. Meanwhile, similar to experiments on the NWPU VHR-10 and ZY-3 image set, the experiments on the MOM-2P images using the proposed method also obtain better results compared to other competitive methods when the scale factor becoming larger. Compared to Table IV with the same upscaling factor, results in Table VI have been improved substantially.

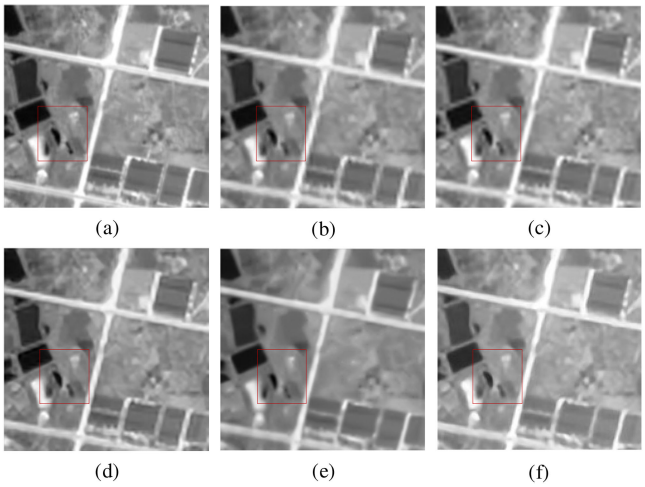


Fig. 7. Reconstructed ZY-3 images with scale factor $s = 3$. (a) Original image. (b) CSRD [26]. (c) BPJDL [27]. (d) SSME [24]. (e) FSRCNN [38]. (f) Proposed method.

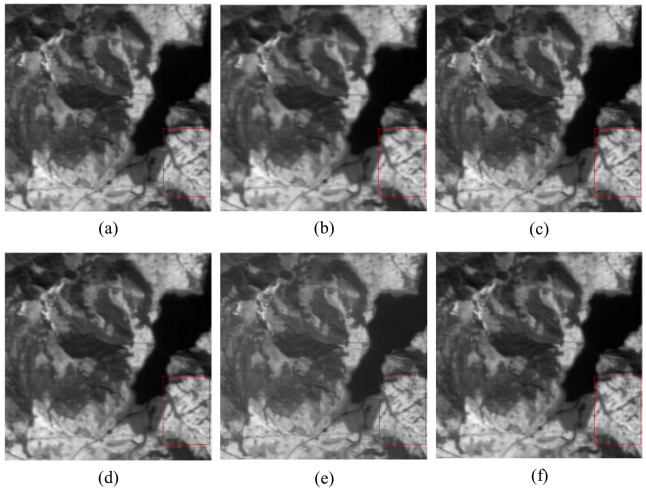


Fig. 8. Reconstructed MOM-2P images with scale factor $s = 2$. (a) Original image. (b) CSRD [26]. (c) BPJDL [27]. (d) SSME [24]. (e) FSRCNN [38]. (f) Proposed approach.

TABLE V
PSNR (dB) AND SSIM RESULTS USING DIFFERENT METHODS ON MOM- 2P IMAGES WITH SCALE FACTOR $s = 2$

	Index	CSRD [26]	BPJDL [27]	SSME [24]	FSRCNN [38]	Proposed
Im11	PSNR	40.04	40.37	38.58	40.52	**40.55**
	SSIM	0.976	0.979	0.955	0.979	**0.980**
Im12	PSNR	38.70	38.92	38.25	**38.93**	38.87
	SSIM	0.971	**0.974**	0.970	0. 973	**0.974**
Im13	PSNR	38.62	39.00	37.15	39.09	**39.11**
	SSIM	0.974	0.977	0.951	0. 977	**0.978**
Im14	PSNR	40.12	40.15	39.78	40.25	**40.36**
	SSIM	0.978	0.979	0.974	0.979	**0.980**
Im15	PSNR	38.99	39.32	37.87	**39.43**	**39.39**
	SSIM	0.969	0.973	0.955	**0.973**	**0.973**
Average	PSNR	39.294	39.552	38.326	39.645	**39.656**
	SSIM	0.9736	0.9764	0.9610	0.9764	**0.9768**

TABLE VI
PSNR (dB) AND SSIM RESULTS USING DIFFERENT METHODS ON MOM- 2P IMAGES WITH SCALE FACTOR $s = 3$

	Index	CSRD [26]	BPJDL [27]	SSME [24]	FSRCNN [38]	Proposed
Im11	PSNR	34.90	35.15	34.18	**35.64**	35.60
	SSIM	0.930	0.935	0.905	0.**939**	**0.939**
Im12	PSNR	33.48	33.78	32.41	33.93	**33.94**
	SSIM	0.910	0.918	0.885	0.920	**0.921**
Im13	PSNR	33.61	33.86	32.76	34.19	**34.21**
	SSIM	0.923	0.930	0.898	0.932	**0.933**
Im14	PSNR	34.78	34.89	33.86	35.11	**35.28**
	SSIM	0.932	0.937	0.892	0.937	**0.940**
Im15	PSNR	34.01	34.23	33.18	34.47	**34.54**
	SSIM	0.910	0.916	0.887	0.920	**0.921**
Average	PSNR	34.156	34.382	33.278	34.670	**34.710**
	SSIM	0.921	0.9272	0.8934	0.9295	**0.9308**

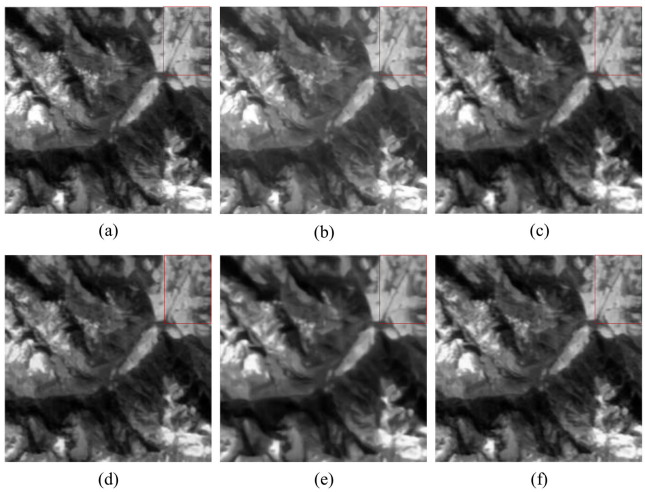


Fig. 9. Reconstructed MOM-2P images with scale factor $s = 3$. (a) Original image. (b) CSRD [26]. (c) BPJDL [27]. (d) SSME [24]. (e) FSRCNN [38]. (f) Proposed approach.

It indicates that images with lower spatial resolution are easier to be reconstructed as there is less high-frequency information.

In Figs. 8 and 9, we show some reconstructed MOM-2P images with $s = 2$ and $s = 3$ respectively. The results demonstrate that the proposed method has also reconstructed the HR images with clear edges and details. Furthermore, with upscaling factor growing larger, the mapping between the LR and HR images becomes more complex, while the proposed method becomes more prominent relative to other compared methods. It indicates that the proposed method greatly benefits from the learned mapping relation with CSAE in sparse domain.

C. Analysis of the Reconstructed Information

In this section, the reconstructed information is further visualized and analyzed. To this end, we first visualize the dictionary

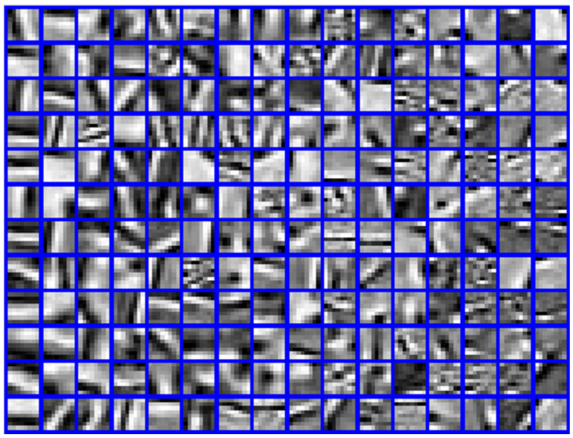


Fig. 10. Redundant dictionary of ZY-3 image.

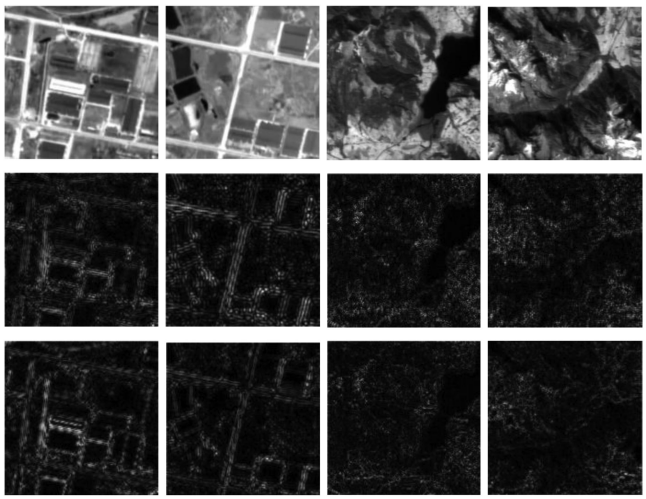


Fig. 11. Reconstructed information of image. They are original image, the reconstructed residuals, and their corresponding ground truth from the first to last raw, respectively.

trained with the sparse decomposition (see Section III-A). However, since each element of the redundant dictionary is a one-dimensional vector, we reshape them into small image patches to have an ease visual effect. Without the loss of generality, the decomposed dictionary of ZY-3 image is provided in Fig. 10. It shows that the redundant dictionary reflects the structural and textural features of the image.

In addition, to directly observe the reconstructed information of the proposed method, we also visualized the reconstructed residuals and their ground truth in Fig. 11 (with scaling for better visual effect). Similar with the redundant dictionary in Fig. 10, the reconstructed residuals in Fig. 11 is also the high-frequency part such as structure and texture. They are just right to supplement the missing high-frequency information of the LR image. Compared to the ground truth, most of the high-frequency information has been reconstructed and the pixels have been accurately reconstructed according to its adjacent pixels.

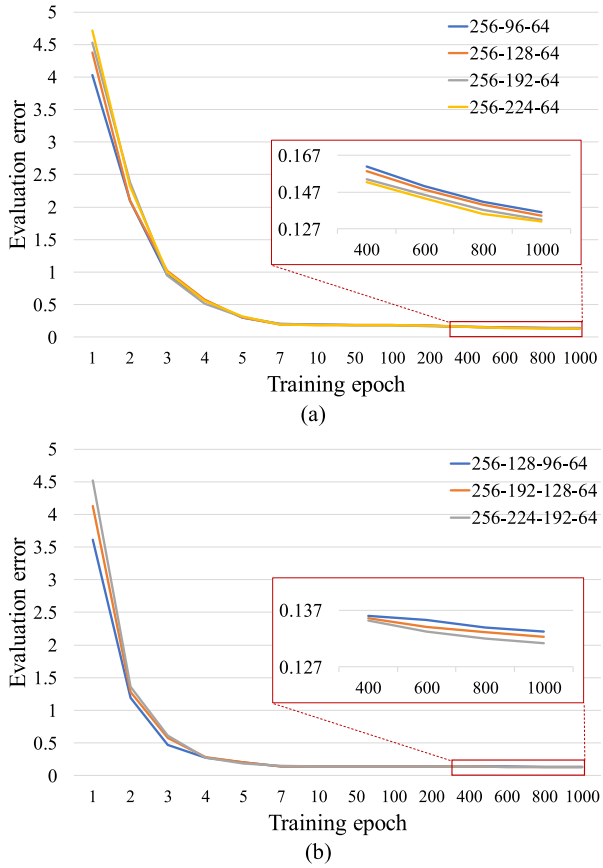


Fig. 12. Relation between the training epoch and the evaluation error with different hidden layers and neuron numbers.

D. Analysis of Parameter Sensitivity

In this section, we conducted more experiments to analyze the parameter sensitivity of the proposed method, mainly including the number of hidden layers and neurons of CSAE and the factor of sparsity constraint for model training. We implemented these experiments on the ZY-3 images with scale factor $s = 3$.

1) Analysis of the Hidden Layers: In the above experiments, only one hidden layer is applied in the proposed CSAE model, and the number of the hidden neurons of the CSAE is set to 192. For convenience, we denote the architecture of the CSAE as 256-192-64, where 256 and 64 are the input and output dimensions, respectively. According to the model parameter settings in Section III-B, we conduct two groups of experiments to explore the number of hidden layers and neurons of the proposed CSAE.

Each experiment is conducted with tenfold cross validation, and the average validation error is reported in Fig. 12. In experiment (a), by comparing the validation results with neuron number 256-96-64, 256-128-64, 256-192-64, and 256-224-64, we can see that the final evaluation error is becoming lower with the increase of the neuron number of the hidden layer. By comparison, 256-192-64 and 256-224-64 demonstrate similar performance but increases the computation cost. In experiment (b), we increase the number of hidden layers of CSAE with

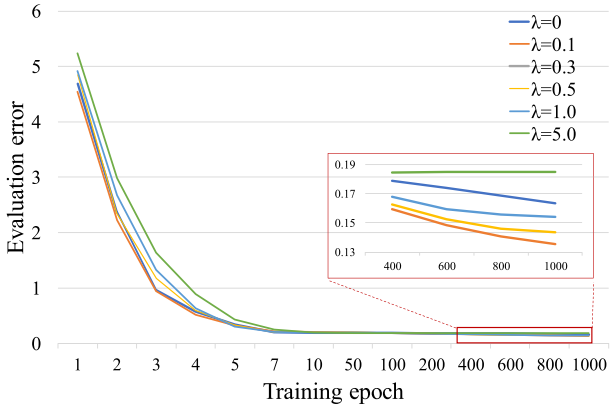


Fig. 13. Influence of different sparsity constraint factors.

TABLE VII
AVERAGE PSNRS (DB) USING DIFFERENT SPARSITY CONSTRAINT FACTORS WITH SCALE FACTOR $s = 3$

Image set	Value of sparsity constraint factor					
	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 1.0$	$\lambda = 5.0$
NWPU VHR-10	25.78	25.85	25.85	25.82	25.80	25.75
ZY-3	31.43	31.49	31.48	31.47	31.46	31.38
MOM-2P	34.60	34.70	34.69	34.67	34.65	34.55

different neuron numbers. We can see that more hidden layers do not result in considerably decreased evaluation error, while the computational complexity can be raised. In fact, a neural network with at least one hidden layer has been proved to has the ability to approximate any continuous function, which also provides a guarantee for our CSAE to approximate the mapping relationship to arbitrary precision.

2) Sensitivity of Sparsity Constraint: We provided more experiments to explore the influence of the sparsity constraint factor λ_l and λ_{hl}, which encourage the sparsity of the reconstructed SCOLRI and SCOR, respectively. Since λ_l and λ_{hl} have the same order of magnitude, we let λ_l and λ_{hl} share the same value as λ for convenience, *i.e.*, $\lambda_l = \lambda_{hl} = \lambda$. Specifically, we applied a series of sparsity constraint factors in CSAE and reported their average performances with ten-fold cross validation (see Fig. 13). Experimental results show that the final validation error tends to a minimum when $\lambda = 0.1$.

In addition, to directly explore the influence of the sparsity constraint on the reconstructed results, we further reported the average PSNR of the reconstructed images on three datasets (see Table VII). Experimental results show that both the too-large and too-small values of the sparsity constraint factor are not conducive to high-quality image SR. Without sparsity constraint $(\text{i.e.}, \lambda = 0)$, the testing PSNR is relatively small due to the overfitting of the model, while large sparsity constraint $(\text{e.g.}, \lambda = 5.0)$ also limits the representation capability of the model.

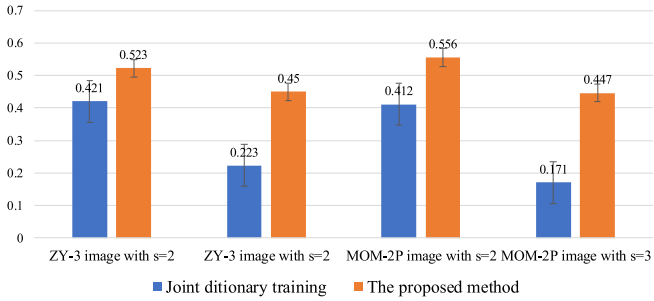


Fig. 14. Correlation coefficients between the true values and predicted sparse coefficients using the proposed method and the joint dictionary training method.

E. Discussion of the Correlation Between Coefficients

To further confirm the efficiency of the proposed method, the correlation coefficients between the predicted SCOR and the ground truth are computed and compared with the joint dictionary training method [20]. The correlation coefficients between the predicted SCOR and the ground truth reflect the consistency of the reconstructed feature information. Larger correlation indicates better reconstruction results and vice versa. For comparison purposes, the corresponding ground truth is generated by decomposing the residuals over its redundant dictionary, and then we estimated it with the proposed method and the joint dictionary training method, respectively. After that, correlation coefficients between the true values and the predicted results are computed and compared.

Fig. 14 shows the correlation coefficients between the predicted SCOR and the ground truth. In general, the proposed method has acquired relatively larger correlation coefficient in each experiment. The mean correlation coefficient of the proposed method reaches 0.494, while the joint dictionary training method is just 0.306, which means the proposed method is more likely to reflect the mapping relation between the sparse coefficients of the LR image and the residuals. In addition, as the upscaling factor grows, the correlation coefficients are getting smaller. With larger upscaling factor, much more high-frequency information is missed, and it is more difficult to construct the mapping relation between the sparse coefficients of the LR image and the residuals. Moreover, with the increasing of the upscaling factor, the gap of correlation between the proposed method and the joint dictionary training method is getting bigger. It means that our method shows far superiority over traditional methods with larger upscaling factors for SR task.

F. Robustness Test

In this section, we further compared the proposed method with FSRCNN [38] on robustness against random Gaussian noise. Specifically, we implemented the robustness test experiment on the ZY-3 image set. The LR images are obtained by downsampling the original images with Bicubic algorithm, and the original images are regarded as the ground truth HR images. During the SR phase, we added random Gaussian noises with a series of variances to the testing LR images to explore the antinoise capacity of the proposed CSAE and FSRCNN [38].

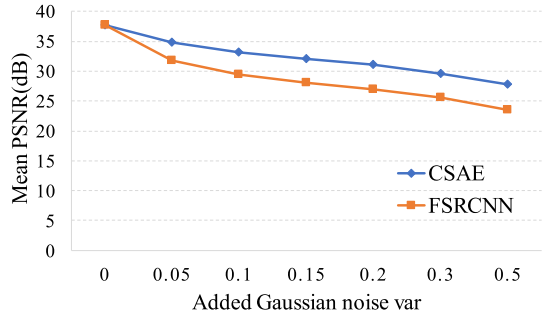


Fig. 15. Robustness against random Gaussian noises with different variances.

The testing results were provided in Fig. 15. It can be seen that, compared to FSRCNN, the proposed method is more robust to noise. In fact, FSRCNN tends to predict the HR image by applying multiple layers of convolution operator. However, the convolution kernels cannot distinguish the noise from the input LR images but regard them as the information of image for the final HR images prediction. Comparatively, the proposed method aims at reconstructing the HR images with the combination of basic structural and textural elements from the redundant dictionary. Therefore, the proposed method is more robust to noise and maintains the edges of the objects well.

V. Conclusion

A novel CSAE is proposed to effectively learn the mapping relation between the LR and HR images for the image SR. Technically, we first decompose the LR and HR images into sparse coefficients and then a CSAE is established to learn the mapping relation between them. The proposed method leverages the feature representation ability of both sparse decomposition and CSAE, and is able to learn the mapping relation between the LR and HR images.

Experimental results on three remote sensing image datasets with different resolutions demonstrate that the proposed method has acquired solid gains than other competitive methods on both visual effects and quantitative indicators. Moreover, at the larger upscaling factors, the proposed method becomes more promising.

Dictionary training for specific information such as edges, textures, and structures is not considered in our work. In future, we will incorporate them into the proposed framework to reconstruct more realistic image profile.

References

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

[2] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[3] Z. Pan *et al.*, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4864–4876, Sep. 2013.

[4] Z. W. Lu, C. D. Wu, D. Y. Chen, Y. C. Qi, and C. P. Wei, "Overview on image super resolution reconstruction," in *Proc. 26th Chin. Control Decis. Conf.*, 2014, pp. 2009–2014.

[5] M. N. Bareja and C. K. Modi, "An effective iterative back projection based single image super-resolution approach," in *Proc. Int. Conf. Commun. Syst. Netw. Technol.*, 2012, pp. 95–99.

[6] S. Hu, S. Zhang, A. Zhang, and S. Chai, "Hyperspectral imagery super-resolution by adaptive POCS and blur metric," *Sensors*, vol. 17, no. 1, 2017, Art. no. 82.

[7] Z. Zhang, X. Wang, J. Ma, and G. Jia, "Super resolution reconstruction of three view remote sensing images based on global weighted POCS algorithm," in *Proc. Int. Conf. Remote Sens., Environ. Transp. Eng.*, 2011, pp. 3615–3618.

[8] Q. Luo, X. Shao, and L. Wang, "Super-resolution imaging in remote sensing," in *Proc. SPIE*, 2015, pp. 175–181.

[9] P. Purkait and B. Chanda, "Super resolution image reconstruction through Bregman iteration using morphologic regularization," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4029–4039, Sep. 2012.

[10] J. Jiang, X. Ma, C. Chen, T. Lu, Z. Wang, and J. Ma, "Single image super-resolution via locally regularized anchored neighborhood regression and nonlocal means," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 15–26, Jan. 2017.

[11] H. Aghighi, J. Trinder, S. Lim, and Y. Tarabalka, "Fully spatially adaptive smoothing parameter estimation for Markov random field super-resolution mapping of remotely sensed images," *Int. J. Remote Sens.*, vol. 36, no. 11, pp. 2851–2879, 2015.

[12] L. K. Tiwari, S. K. Sinha, S. Saran, V. A. Tolpekin, and P. L. N. Raju, "Markov random field-based method for super-resolution mapping of forest encroachment from remotely sensed ASTER image," *Geocarto Int.*, vol. 31, no. 4, pp. 428–445, 2016.

[13] Y. Zhao, J. Yang, and J. C. W. Chan, "Hyperspectral imagery super-resolution by spatial-spectral joint nonlocal similarity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2671–2679, Jun. 2014.

[14] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2014.

[15] W. Wang, C. Ren, X. He, H. Chen, and L. Qing, "Video super-resolution via residual learning," *IEEE Access*, vol. 6, pp. 23767–23777, 2018.

[16] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," 2016, arXiv: 1609.04802.

[17] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.

[18] K. Zeng, J. Yu, R. Wang, C. Li, and D. Tao, "Coupled deep autoencoder for single image super-resolution," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 27–37, Jan. 2017.

[19] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

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

[21] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1382–1394, Apr. 2013.

[22] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3991–4002, Nov. 2017.

[23] S. Gou, S. Liu, S. Yang, and L. Jiao, "Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4784–4792, Dec. 2014.

[24] W. Xinlei and L. Naifeng, "Super-resolution of remote sensing images via sparse structural manifold embedding," *Neurocomputing*, vol. 173, pp. 1402–1411, 2016.

[25] C. He, L. Liu, L. Xu, M. Liu, and M. Liao, "Learning based compressed sensing for SAR image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1272–1281, Aug. 2012.

[26] Y. Sun, G. Gu, X. Sui, Y. Liu, and C. Yang, "Single image super-resolution using compressive sensing with a redundant dictionary," *IEEE Photon. J.*, vol. 7, no. 2, Apr. 2015, Art. no. 6900411.

[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 Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 345–352.

[28] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2569–2582, Jun. 2014.

[29] F. Yeganli, M. Nazzal, M. Unal, and H. Ozkaramanli, "Image super-resolution via sparse representation over multiple learned dictionaries based on edge sharpness," *Signal, Image Video Process.*, vol. 10, no. 3, pp. 535–542, 2016.
[30] Y. Zhang, J. Liu, W. Yang, and Z. Guo, "Image super-resolution based on structure-modulated sparse representation," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2797–2810, Sep. 2015.
[31] 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.
[32] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3862–3871.
[33] Z. Zhang, A. Liu, and Q. Lei, "Image super-resolution reconstruction via RBM-based joint dictionary learning and sparse representation," *Proc. SPIE*, vol. 9815, 2015, Art. no. 9815287.
[34] B. Hou, K. Zhou, and L. Jiao, "Adaptive super-resolution for remote sensing images based on sparse representation with global joint dictionary model," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2312–2327, Apr. 2018.
[35] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[36] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[37] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016.
[38] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.

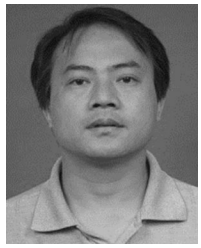
Zhenfeng Shao received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2004, working with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS).

Since 2009, he has been a Full Professor with LIESMARS, Wuhan University. He has authored or coauthored more than 50 peer-reviewed articles in international journals. His research interests include high-resolution image processing, pattern recognition, and urban remote sensing applications.

Dr. Shao was the recipient of the Talbert Abrams Award for the Best Paper in image matching from the American Society for Photogrammetry and Remote Sensing in 2014 and the New Century Excellent Talents in University from the Ministry of Education of China in 2012. Since 2019, he has been an Associate Editor for the *Photogrammetric Engineering and Remote Sensing*, specializing in smart cities, photogrammetry, and change detection.

Lei Wang received the B.Sc. degree in 2015 from the School of Mathematics and Statistics, Wuhan University, Wuhan, China, where he is currently working toward the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing.

His research interests include pattern recognition and machine learning.



Zhongyuan Wang received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008.

He is currently a Professor with Computer School, Wuhan University. He has directed three projects from NSFC and has authored or coauthored more than 80 papers in distinguished international conferences and journals. His research interests include video compression, image processing, and multimedia big data analytics.

Juan Deng received the M.Eng. degree in geomatics engineering in 2017 from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, where she is currently working toward the Ph.D. degree with the School of Electronic Information.

Her research interests include pattern recognition and deep learning.