Convolutional Sparse Coding using Wavelets for Single Image Super-Resolution

AWAIS AHMED¹, SHE KUN¹, RAHEEL AHMED MEMON², JUNAID AHMED³, GETNET TEFERA¹,
¹School of Information and Software Engineering University of Electronic Science and Technology, Chengdu, China; e-mail: engr.awais88@yahoo.com; kunshe@126.com; getmary21@gmail.com
²Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan; e-mail: raheelmemon@iba-suk.edu.pk
³Department of Electrical Engineering, Sukkur IBA University, Sukkur 65200, Pakistan; e-mail: j.bhatti@iba-suk.edu.pk
Correspondence: Awaiz Ahmed (engr.awais88@yahoo.com); She Kun (kunshe@126.com)

ABSTRACT In this paper, we propose the convolutional sparse coding based model in the wavelet domain for the task of single image super-resolution (SISR). The conventional sparse coding based approaches work on overlapping image patches and use the dictionary atoms to sparse code an image patch. Further, at the final stage, an overlap-add mechanism is used to get the final high-resolution image estimate. However, these algorithms fail to take into account the consistency present in the overlapping patches which limits their performance. We propose the use of wavelet integrated convolutional sparse coding approach where instead of dictionary atoms we utilize the convolution summations between the learned filters and their mappings for sparse representation based SISR. The use of wavelets is proposed owing to their unique directional and compact features. A pair of filters are learned along with a mapping function for each wavelet sub-band to exploit the consistency among patches. The proposed wavelet integrated convolutional sparse coding model helps capture useful contextual information. The proposed model is evaluated on publicly available datasets for different scale-up parameters. To show the efficacy of the proposed model we compare it with recent state-of-the-art algorithms. The visual results along with the quantitative ones indicate that the proposed model performs well for the tasks of super-resolution.

INDEX TERMS Super-resolution, coupled filters learning, mapping learning, sparse coding.

I. INTRODUCTION
The task in Single Image Super-Resolution (SISR) is to estimate the High Resolution (HR) image from a Low Resolution (LR) observation. The problem of SISR is quite old but still active and challenging research problem mainly because of its ill-posed nature and high demand. Generally, for the task of SISR, we model the output of a degradation model as;

\[ l = (h \odot d) \downarrow s \quad (1) \]

where \( h \odot d \) is the convolution operation between the HR image \( h \) and the blur operator \( d \), \( \downarrow s \) is the down sampling with the scale \( s \).
The SISR algorithms can be generally classified into three types, i.e., interpolation algorithms, model-oriented optimization algorithms, and discriminative based learning algorithms. The interpolation algorithms such as Bicubic, Bilinear and Nearest Neighbor [1],[2],[3] are very easy to implement and quite efficient however, their performance is quite limited. The second class of algorithms the optimization based algorithms usually employ a specific prior in the optimization model such as sparsity based prior [4], self-similarity based prior [5] and denoise prior [6]. These algorithms are quite flexible and can achieve high-quality results. The main problem with these algorithms is the time consumption. In recent years, a lot of focus has been put on the discriminative based learning algorithms. This is due to the circumstance that these algorithms employ the Convolution Neural Network (CNN) architecture and possess their unique properties [7]. In [8] the
authors proposed a unified spatial-temporal-spectral framework based on a deep convolutional neural network (STS-CNN). It is a learning based model for focusing on the rebuilding of remote sensing imagery contaminated through dead pixels and thick cloud. There proposed model can utilize multiple data (spatial, spectral and temporal) input for there (STS-CNN) for the different task of reconstruction. In [9] the authors proposed a joint spatial-spectral residual network with multiscale characteristics extraction to improve the noise free hyperspectral imagery (HIS). They utilized mutually spatial structure and adjacent interrelated spectra are instantaneously proposed for extraction characteristics and representation.

In this article, we propose a wavelet domain based algorithm utilizing the Stationary Wavelet Transform (SWT) [10],[11]. The SWT is preferred over the DWT owing to its useful up-sampling decomposition. The wavelet sub-band images are significantly sparse. Moreover, wavelet images have unique properties such as sparsity, compactness, directionality, and redundancy. We propose the convolution sparse coding based model using wavelets for the task of SISR. We learn a set of HR-LR filters along with the HR-LR mappings in the wavelet domain. We learn the LR filters for the wavelet sub-band images at level-2 of wavelet decomposition and find their feature maps. We learn the mappings between level-1 HR and level-2 LR sub-band images to predict the level-1 HR feature maps from the level-2 LR. Finally, we learn the HR filters to estimate the desired level-1 HR wavelet sub-band images using the level-1 HR feature maps by simple convolutions. As the wavelet sub-band images have a very useful property of redundancy, we can exploit this property by convolution-based sparse HR and LR filters and mapping learning. The comparison of our proposed algorithm is carried out with modern state-of-the-art algorithms on peak-signal-to-noise-ratio (PSNR), information fidelity criterion (IFC), running time, and structural similarity index measure (SSIM) quantitative parameters on different scale factors. For testing, all the algorithms are compared over the publicly available datasets of “Set5”, “Set14”, “BSD100” and “Urban100”, these are publicly available. The quantitative and qualitative results specified that our proposed algorithm achieves improved results in reconstructing sharp and directional fine features for the problem of SISR. The key points of the proposed algorithm can be summarized as;

- The proposed model utilizes the convolutional sparse coding integrated with wavelet analysis for the task of SISR.
- The inconsistency problem of overlapping patches is solved by representing the whole image as a summation of convolutions.
- The useful properties of directionality and redundancy of SWT are exploited using the convolutional sparse coding.
- The convolutional sparse coding model is improved using undecimated stationary wavelet transform. As the wavelets induce stationary wavelet transform.

II. RELATED WORK

The first CNN based model for performing the task of SISR was proposed by Dong et al. [10] called super-resolution using the convolutional neural network (SRCNN). This model has a three-layered network architecture. In their extended work in [11], they investigated the depth of the network and modeled the high depth CNN networks. They showed that by increasing the depth, limits the performance of the CNN based super-resolver. This problem was solved by Kim et al. [12], they proposed a very deep super-resolution (VDSR) network architecture for performing the task of SISR. They proposed the residual image learning then to boost the speed they used adjustable gradient clipping. They also showed that residual learning can develop a single model that can work for multiple scale degradations. These algorithms use the input LR image which is bicubic interpolated as an input to the network. This task increases the computational cost and limits their performance. Recently, a lot of neural network based models are proposed for execution the task of SISR. In [13] proposed a convolutional sparse coding (CSC) based algorithm for the task of SISR. Here authors learn a set of HR and LR filters along with mapping functions for the whole image instead of working on patches. In [14] authors propose a novel sparse coding network (SCN) for performing the task of SISR. They utilize the framework of sparse coding and develop a deep neural network architecture. They utilize LISTA as a sub-network for sparse coding with recurrent stages. In [15] authors propose an example learning based algorithm utilizing the residual weighted random forest and non-local similar structures for the task of SISR. Furthermore, k-means clustering is applied to exploit the non-local similar structures. In [16], authors propose a feed-forward network based on denoising convolutional neural network (DnCNN). Here, authors utilize the residual learning and batch normalization operations for their model. They learn a single model for different image processing tasks such
as denoising, deblocking and SISR. In [17], authors propose a more fast and robust extension to SRCNN [10] called fast super-resolution convolutional neural network (FSRCNN). They achieve this in a three-fold approach. First, a deconvolution layer is introduced. Second, the interpolation is avoided for LR to HR. Third, mapping layers are modified by reducing their size and increasing their number. In [18], the authors propose a neural network model for super-resolution with multiple degradations (SRMD). They utilize a principal component analysis based dimensionality stretching to learn a single network for denoising and super-resolution task. Moreover, their network can handle multiple spatially verifying degradations. In [19], authors propose a deep & compact information distillation network (IDN) for the task of SISR. Group convolutions and small filter sizes are used for learning. Their algorithms work in three steps of block feature extraction, distillation & finally reconstruction of the blocks.

Another category of SISR based algorithms can be the wavelet-based algorithm. In [20], Nazzal et al. proposed a dictionary learning based algorithm for the task of SISR. They utilize the unique properties of the Discrete Wavelet Transform (DWT) for the task of SISR. In [21], Ahmed et al. proposed a coupled dictionary learning algorithm in the wavelet domain for the task of SISR. In [22], extended work is proposed using the Dual-Tree Complex Wavelet Transform (DT-CWT) plus a coupled dictionary and mapping learning for performing the task of SISR. In [23], Selen et al. proposed a novel model in a wavelet domain employing dictionary learning for the execution of the task of SISR. In [24], convolution neural network in wavelet domain for super-resolution (CNNWSR) is proposed. Here, the authors proposed a wavelet domain based CNN model for performing the task of SISR. The authors argue that as the wavelet sub-band images and corresponding approximations have a strong intra-scale dependency which can be exploited for the execution of the SISR. They utilize the DWT for wavelet domain analysis and synthesis. In [25] authors proposed a model called Deep Wavelet Super-Resolution (DWSR) in their model they merged the data into low and high frequency sub-bands in the wavelet domain by using a deep CNN architecture. Furthermore, they build maximum residual networks that is proper for the wavelet coefficients because of their sparsity promoting nature. In [26] author’s key direction was that to progress a new CNN structure that overcomes the limitation of the state-of-the-art CNN techniques. They proposed a principle network by utilizing the manifold simplification. Secondly employing the current computational topology tool called the persistent homology, they show that the current residual learning is a distinct case and after simplification, they proposed a wavelet transform to simplify topological constructions of input or label manifolds. In [27] the authors present multi-level wavelet CNN (MWCNN) model for improved compromise between performance and competence. By the modification of U-NET architecture, in this work they used discrete wavelet transform (DWT) to reduce the size of characteristics in constructing subnetwork for the task of SISR.

III. PROPOSED METHODOLOGY

A. STATIONARY WAVELET TRANSFORM

The wavelet transform has many unique properties such as multi-scale analysis which makes it suitable to be utilized in the image processing applications. Unlike the discrete wavelet transform (DWT), the stationary wavelet transform (SWT) is an up-sampling method. Therefore, after applying the SWT the size of the approximation image and other wavelet sub-band images does not change.

In comparison with DWT based SISR algorithms [23-27], no down-sampling is made in SWT due to which it has certain advantages;

- We can preserve more details using the SWT analysis. The SWT analysis uses the up-sampling process instead of the decimation so, the size of the images after the analysis process does not change. This, in turn, helps preserve more details and the translation invariance of the wavelets is kept in check by the over-complete representation [28].
- The difference in the value range of the wavelet coefficients increases with an increase in levels of decomposition. This helps the HR-LR filters and mapping functions to efficiently encode and decode images on the different wavelet decomposition levels. However, due to the redundancy property of the wavelets, the position of the coefficient values is invariant to the scale-up parameters.

IV. CONVOLUTIONAL SPARSE CODING

The method of sparse coding despite being used in most of the image processing applications has some shortcomings [4],[29],[20]. For example, the scalability of $l_0$ or $l_1$ norm in such algorithms is not good enough to be used in large scale applications which are overcome by neural network and other algorithms utilizing full image up-sampling instead of based approach [10], [12], [13], [14]. The second shortcoming can be that as most of the sparse coding problems divide the input image into overlapping patches and then apply the algorithm to reduce the computational burden on the algorithm [20], [21], [22]. However, they ignore the consistency present in those overlapping patches. To alleviate this problem aggregation and averaging
strategies are employed to some extent. However, still, there is room for improvement.

To tackle this problem of inconsistency, authors in [30] proposed an algorithm to sparse code the complete images not in patches using the convolution summations. Instead of representing it by a dictionary atom and the corresponding sparse coding vector, their sparse coding model uses the convolution summation between the convolution filters and feature maps:

$$\min_M \|X - \sum_{i=1}^{N} h_i \odot M_i \|_F^2 + \lambda \sum_{i=1}^{N} \|M_i\|_1$$  \hspace{1cm} (2)$$

where $X$ is an input image with $m \times n$ dimensions, $h_{i=1,2,...,N}$ represents the $w \times w$ filters $\lambda$ is a regularizing parameter and $N$ is the number of filters and $M_i$ represents the feature maps having sizes $(m + w - 1) \times (n + w - 1)$. By using this strategy of summation of feature maps and filters, the inconsistency problem is tackled. It is also worth noting that the size of the feature maps $M_i$ is almost same as $X$.

Despite being a good way to tackle the inconsistency problem, the convolution sparse coding problem suffers from poor optimization. Zeiler et al. [31] proposed the conjugate gradient (CG) technique to solve that problem. Bristow et al. [32] proposed a block circulant with circulant block (BCCB) algorithm in the Fourier domain to solve this problem. Wolberg et al. [33] further proposed an improved algorithm which is based on the alternating direction method of multipliers (ADMM) for the optimization of this problem.

V. PROPOSED MODEL

The convolutional sparse coding utilizes the convolution summation for up-sampling the LR image instead of dictionary learning. Moreover, the patches-based approach is avoided and the whole image is processed which helps capture the image features as a whole. Furthermore, to improve the results and decrease the computational cost the sparsity, and directionality is induced by using the SWT. In comparison with the spatial domain algorithm, the wavelet domain offers significant improvements in capturing the directional features of the images with better resolution. Further, with the inherent sparsity property of wavelets, the filter sizes of the proposed method are reduced in comparison with the spatial domain which results in reduced computational cost.

In the proposed model we have utilized the SWT wavelet transform which has the same approximate properties as the bicubic degradation. The level-1 sub-band images are considered the HR images and corresponding the level-2, level-3 and level-4 are considered LR images depending on the scale-up being performed. Once the HR and LR images performed. Here for the training we utilize the 3/2 ratio of HR/LR filter as suggested by the [13]. The spatial domain CSC [13] algorithm uses LR filters as the ratio of HR/LR filters to achieve good results with reasonable computational cost. For our experiment we use 200 LR filters with the ratio of 3/2 for HR/LR filters. If the filters number is increased the results also improve [13] but the computational cost also increases. So, an optimum is chosen based on experimental analysis. The training is done in three parts. First LR filters are learned using LR images. Second, LR and HR features are extracted from HR and LR images. Third, a joint HR filters and mappings are learned between HR and LR features using HR images and extracted features. In the testing stage, given the LR image, the learned LR filters are used to extract the LR feature maps. Then, learned mapping is used to extract approximate HR feature maps for LR feature maps. The HR image is estimated by using the HR feature maps. The flow chart of algorithm is given in Fig.1.

VI. THE TRAINING ALGORITHM

In most of the dictionary learning based super-resolution algorithms a pair of HR and LR dictionaries along with the mapping functions are used to model the single optimization problem and dictionaries are trained jointly on HR and LR patch pairs. However in the testing phase where test HR image is not given and sparse coding vectors in the training and testing are not necessarily the same and thus they induce inconsistency. To tackle this problem several algorithms are developed [28] which solve a bi-level optimization problem. In recent years separate dictionary training is also encouraged and is reported to give nice results [34]. Here, authors learn an LR dictionary from the LR training data and then the HR dictionary is learned to get the HR image using the corresponding LR sparse coefficients. Here, we extend this method in our proposed approach.

A. LEARNING LR FILTERS

Let us consider we have generated the HR and LR training images by using the 2-level SWT decomposition and denote them by; $X^s = X_1^s, X_2^s, \ldots, X_k^s$ as the HR images set and $Y^s = Y_1^s, Y_2^s, \ldots, Y_k^s$ as the LR images set. It should be noted here that as $X^s$ and $Y^s$ contains the set of all the three sub-band images i.e. horizontal, vertical and diagonal denoted by subscript $s$. For the sake of simplicity, we omit subscript $k$. 
Given the set of LR sub-band images $Y^s$ we learn the LR filters and feature maps [30].

$$\min_{M^s, h^s} \|Y^s - \sum_{i=1}^{N} h^s_i \odot M^s_i \|_2^2 + \lambda \sum_{i=1}^{N} \|M^s_i \|_F$$

where $h^s_i$ are the N LR filters and $M^s_i$ is the sparse feature map for the $i^{th}$ filter. We use the alternate optimization for the $M^s$ and $h^s$ sub-problems. For the sub-problem in Eq.3 for $M^s$ which is a standard convolution sparse coding problem we use the algorithm of [33]. For the sub-problem of $h^s_i$ we reformulate the Eq. 3 as:

$$h^s_i = \arg \min_{h^s_i} \|Y^s - \sum_{i=1}^{N} h^s_i \odot M^s_i \|_2^2 \quad \text{s.t} \quad \|h^s_i \|_F^2 \leq 1$$

(4)

This can be solved by the stochastic average ADMM [35] algorithm.

### B. HIGH RESOLUTION FILTER AND JOINT MAPPING LEARNING

Afterward learning the LR filters and feature maps we learn the HR filters and mapping functions between the LR features maps and HR images. In almost all the SISR algorithms LR images are interpolated to be the same size as the HR. However in our case, as we are using the SWT this step is skipped. The mappings between the HR and LR can be given by:

$$M^h_j(x, y) = g[M^l_1(x, y), M^l_2(x, y), ..., M^l_N(x, y); w^s]$$

(5)

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.
parameter of mapping function \(g(\cdot)\). Here for \(M_h^{I_s}(x, y')\) with the \(\text{mod}(x', k) \neq 0\) or \(\text{mod}(y', k) \neq 0\) we set \(M_h^{I_s}(x, y') = 0\).

The function \(g(\cdot)\) should be able to obtain sparse output given the sparse input:

\[
M_h^{I_s}(x, y) = g\left(\left(\sum_{j=1}^{M_a^{I_s}} M_{a_s}^{I_s}(x, y)\right); w_f^j\right) = w_f^j \sum_{m=1}^{a_s} m_a^{I_s}(x, y), s.t \ w_f^j \geq 0, \|w_f^j\| = 1
\]

where \(m_a^{I_s}(x, y)\) are the coefficients for the point \((x, y)\) of the LR, \(a\) feature maps and \(w_f^j\) is the transformation vector of the HR feature maps \(M_h^{I_s}\). We let \(w_f^j \geq 0\) and \(\|w_f^j\| = 1\) to certify the sparsity of \(w_f^j\).

The non-negative simplex constraint present in Eq. 6 is stronger than some of the \(l_1\) norm based approaches. The joint HR filter and mapping learning model can be formulated as:

\[
h^{I_s, w_s} = \min_{h^{I_s, w_s}, w_s} \|X - \sum_{j=1}^{M} a_j^{I_s} \otimes h_w^s \|_F^2, s.t \ \|h_w^s\|_1 \leq e_j; w_f^j \geq 0, \|w_f^j\| = 1
\]

where \(e\) is the scalar to constraint the energy of HR filters. Here again, we optimize the objective function by alternately updating the filter \(h^{I_s, w_s}\) and mapping parameter \(w_s^j\). For the filter part, we solve it by the algorithm of [33] and for the \(w_s^j\) we solve the following optimization problem:

\[
w_s^j = \arg \min_{w_s^j} \|X - \sum_{l=1}^{M_a^{I_s}} M_{a_s}^{I_s} \otimes h_w^s \|_F^2, s.t \ w_f^j \geq 0, \|w_f^j\| = 1
\]

we solve the problem in Eq. 8 by the algorithm of [35].

**VII. THE TESTING ALGORITHM**

Given the LR filters \(h^{I_s}\), HR filters \(h^{I_s}\) and the mapping function \(w_s^j\). For a selected LR image we first perform the level-1 wavelet decomposition to get the LR sub-band images. We extract the LR sparse feature maps \(M^{I_s}\) using the LR filters. Next, we estimate the HR feature maps using the learned mappings and LR feature maps by \(M^{h_s} = gM^{I_s}; w_s^j\). Finally, the high resolution sub-bands are formulated by convolutions between the HR feature maps and corresponding HR filters.

\[
X^s = \sum_{j=1}^{M} a_j^{I_s} \otimes M_{a_s}^{I_s}
\]

We then perform the 1-level inverse wavelet transform to get the HR image estimate.

**VIII. EXPERIMENT AND RESULTS**

Now we present the details about the experiments and parameters setting of our algorithm. Also, a comparative analysis is presented with other state-of-the-art algorithms. The experimental setup in our case is similar to that of Yang et al. [4]. For training, we utilized the same set of 91 images which is given by Yang et al. [4]. In the wavelet domain for each sub-band (horizontal, vertical and diagonal) we sample approximately 1000, 62 \(\times\) 62 blocks for training and to avoid the boundary condition we pad 8 pixels in each block.

For the parameters setting the most important parameters are LR and HR filters numbers \(N\) and \(M\). The algorithm of CSC [13] uses the \(3/2\) ratio between the HR and LR filter numbers and train around 800 LR filters. For our case, as wavelets are already sparse we train a small number of 200 LR filters with the same HR/LR filter ratio of \(3/2\). In this way, we train 3 sets of HR/LR filters with smaller sizes as compared to the algorithm of CSC [13]. Further, we set the LR and HR filter sizes as \(5 \times 5\). The regularization parameter \(\lambda\) is set to 0.02. The energy constraint parameter for the filters is set to 4, 9 and 12 for scale-up parameters of 2, 3 and 4.

The proposed algorithm is compared with the recent state-of-the-art algorithms of SRMD [18], the algorithm of IDN [19], the algorithm of DnCNN [16], the algorithm of FSRCNN [17], very deep convolutional network (VDSR) [12], convolutional sparse coding (CSC) [13], super-resolution convolutional neural network (SRCNN) [10], and Bicubic technique. All the algorithms are simulated for the scale-up parameter of 2, 3 and 4 over the publicly available data set of ’Set-5′, ’Set-14′, ’BDS100′ and ’Urban-100′. The algorithms under comparison were downloaded from the author’s website and run on Matlab-2017b software with PC configurations of i7 3.2GHz CPU with 8GB RAM. For the computational cost of the proposed algorithm which is mainly dominated by the CSC [13] sparse coding step. The computational cost of our algorithm can be given as \(O(3KN\log N)\). Where \(3K\) are the filters and \(N\) are the pixel number on the images. As compared to the CSC, the proposed algorithm has almost similar computational cost as the filter numbers are reduced. However, where CSC uses a single large filter number, we use a smaller number of filters as in IDN [19] but 3 filters for each wavelet sub-band.

Table. 1 gives a comparative analysis based on PSNR and SSIM for the proposed algorithm. It can be seen from the Table. 1 that on average the proposed algorithm gives better PSNR and SSIM performance. Table. 2 shows the comparative results based on information fidelity criterion (IFC) and time consumption. The highlighted values of algorithms...
show the best results. It can be observed from the Table. 2 that the proposed algorithms give reasonable better results in comparison with state-of-the-art algorithms. The proposed algorithm is better than CSC [13] and DnCNN [16] in terms of computational cost. The visual results are given from Fig. 2 to Fig. 7. Fig. 2 gives the visual comparison on scale-up parameter 2 for the 'Barbara' image. Looking at the figure, it can be seen that the proposed algorithm is able to perform well.

By inspecting the zoomed part of the images, the results of the proposed algorithm are on par with the algorithms of SRMD [18], IDN [19], and DnCNN [16]. The results from other algorithms are still blurry and have ringing artifacts.

The Fig. 3 shows the visual comparison on the scale-up parameter 3 for the 'Man' image from the 'BSD100' data set. Looking at the zoomed parts,
FIGURE 4. Visual comparison of City image for scale-up 4: (a) Bicubic; (b) DnCNN [16]; (c) FSRCNN [17]; (d) IDN [19]; (e) Original; (f) CSC [13]; (g) SRCNN [10]; (h) SRMD [18]; (i) VDSR [12]; (i) Proposed; from Urban-100 data set.

FIGURE 5. Visual comparison of Foreman for scale-up 2: (a) N-CNN [26]; (b) MWCNN [27]; (c) Proposed; (d) Original; from Set-14 dataset.

FIGURE 6. Visual comparison of Worker image for scale-up 3: (a) N-CNN[26]; (b) MWCNN [27]; (c) Proposed; (d) Original; from BSD100 dataset.

FIGURE 7. Visual comparison of Worker image for scale-up 3: (a) N-CNN[26]; (b) MWCNN [27]; (c) Proposed; (d) Original; from Urban100 dataset.
<table>
<thead>
<tr>
<th>Method</th>
<th>Set 5</th>
<th>Set 14</th>
<th>BSD100</th>
<th>Urban100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bicubic</td>
<td>33.64</td>
<td>30.39</td>
<td>28.42</td>
<td>30.22</td>
</tr>
<tr>
<td></td>
<td>0.929</td>
<td>0.866</td>
<td>0.810</td>
<td>0.868</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.907</td>
<td>0.862</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>CSC</td>
<td>36.52</td>
<td>32.60</td>
<td>30.39</td>
<td>32.42</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.907</td>
<td>0.862</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>VDSR</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>DnCNN</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>FSRCNN</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>SRMD</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>IDN</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>MWCNN</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>N-CNN</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
<tr>
<td>Proposed</td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>0.953</td>
<td>0.922</td>
<td>0.885</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>37.56</td>
<td>33.67</td>
<td>31.35</td>
<td>33.02</td>
</tr>
<tr>
<td></td>
<td>37.58</td>
<td>33.75</td>
<td>31.40</td>
<td>33.03</td>
</tr>
<tr>
<td></td>
<td>36.99</td>
<td>33.15</td>
<td>30.71</td>
<td>32.73</td>
</tr>
</tbody>
</table>

**TABLE 2** AVERAGE IFC (TOP) AND TIME (BOTTOM) QUANTITATIVE COMPARISON RESULT FOR DIFFERENT ALGORITHMS.

<table>
<thead>
<tr>
<th>Method</th>
<th>Set 5</th>
<th>Set 14</th>
<th>BSD100</th>
<th>Urban100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Bicubic</td>
<td>6.082</td>
<td>3.579</td>
<td>2.328</td>
<td>6.104</td>
</tr>
<tr>
<td></td>
<td>7.023</td>
<td>4.212</td>
<td>2.893</td>
<td>6.845</td>
</tr>
<tr>
<td></td>
<td>8.047</td>
<td>4.963</td>
<td>2.986</td>
<td>7.730</td>
</tr>
<tr>
<td></td>
<td>17.48</td>
<td>17.39</td>
<td>16.74</td>
<td>23.98</td>
</tr>
<tr>
<td>CSC</td>
<td>25.91</td>
<td>25.11</td>
<td>25.69</td>
<td>25.76</td>
</tr>
<tr>
<td></td>
<td>7.543</td>
<td>4.625</td>
<td>3.101</td>
<td>7.341</td>
</tr>
<tr>
<td></td>
<td>6.206</td>
<td>2.860</td>
<td>1.916</td>
<td>5.288</td>
</tr>
<tr>
<td></td>
<td>8.017</td>
<td>4.963</td>
<td>3.407</td>
<td>7.796</td>
</tr>
<tr>
<td>VDSR</td>
<td>4.787</td>
<td>3.709</td>
<td>4.169</td>
<td>8.410</td>
</tr>
<tr>
<td></td>
<td>8.169</td>
<td>4.987</td>
<td>3.450</td>
<td>7.858</td>
</tr>
<tr>
<td></td>
<td>17.48</td>
<td>17.39</td>
<td>16.74</td>
<td>23.98</td>
</tr>
<tr>
<td>DnCNN</td>
<td>25.91</td>
<td>25.11</td>
<td>25.69</td>
<td>25.76</td>
</tr>
<tr>
<td></td>
<td>7.543</td>
<td>4.625</td>
<td>3.101</td>
<td>7.341</td>
</tr>
<tr>
<td></td>
<td>6.206</td>
<td>2.860</td>
<td>1.916</td>
<td>5.288</td>
</tr>
<tr>
<td></td>
<td>8.017</td>
<td>4.963</td>
<td>3.407</td>
<td>7.796</td>
</tr>
<tr>
<td>FSRCNN</td>
<td>8.047</td>
<td>4.963</td>
<td>2.986</td>
<td>7.730</td>
</tr>
<tr>
<td></td>
<td>6.206</td>
<td>2.860</td>
<td>1.916</td>
<td>5.288</td>
</tr>
<tr>
<td></td>
<td>8.017</td>
<td>4.963</td>
<td>3.407</td>
<td>7.796</td>
</tr>
<tr>
<td>SRMD</td>
<td>3.585</td>
<td>2.106</td>
<td>0.685</td>
<td>4.065</td>
</tr>
<tr>
<td></td>
<td>9.251</td>
<td>5.619</td>
<td>3.820</td>
<td>8.837</td>
</tr>
<tr>
<td></td>
<td>3.299</td>
<td>2.492</td>
<td>1.519</td>
<td>5.334</td>
</tr>
<tr>
<td></td>
<td>8.999</td>
<td>5.563</td>
<td>3.948</td>
<td>8.732</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.596</td>
<td>0.457</td>
<td>0.455</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>7.678</td>
<td>4.812</td>
<td>3.365</td>
<td>7.176</td>
</tr>
<tr>
<td></td>
<td>3.595</td>
<td>2.287</td>
<td>2.657</td>
<td>3.622</td>
</tr>
<tr>
<td></td>
<td>8.797</td>
<td>5.639</td>
<td>4.497</td>
<td>8.458</td>
</tr>
<tr>
<td></td>
<td>15.49</td>
<td>15.47</td>
<td>15.44</td>
<td>15.53</td>
</tr>
</tbody>
</table>

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.
it can be seen that the proposed algorithm performs better in terms of texture recovery. Comparing the results with the SRMD [18], DnCNN [16] and IDN [19] algorithms the proposed algorithm has good resolution and performance. The visual results also validate the quantitative results.

Fig. 4 shows the result from the more challenging data set of 'Urban-100'. Here the scale-up parameter is 4. Observing the sub-figures, it can be seen that almost all the algorithms fail to recover the high-frequency texture from the tall image buildings. From the zoomed parts of the sub-images, it can be observed that the proposed algorithm has reasonable results in comparison with the algorithms of DnCNN [16], SRMD [18], and IDN [19].

Fig. 5 shows the visual comparison on the scale-up parameter 2 for the ‘foreman’ image which is taken from Set-14 data set. Fig. 6 shows the visual comparison on scale-up 3. This image is taken from BSD100 dataset. Fig. 7 shows a building image on scale-up parameter 4, this image is taken from Urban100 dataset. From these results, it can be seen that the proposed algorithm has better reconstruction ability than the algorithms of N-CNN [26] and MWCNN [27] on the compared set of images.

IX. CONCLUSION
In this paper, a new model is proposed for the task of SISR. The proposed model utilizes the convolutional sparse coding approach along with the wavelet analysis for the task of SISR. The inconsistency problem of overlapping patches is solved by representing the whole image as a summation of convolutions. The useful properties of directionality and redundancy of SWT are exploited using the convolutional sparse coding. The results based on PSNR and SSIM indicate that the proposed model gives good results for the SISR problem. The visual results also validate the PSNR and SSIM results.

The proposed algorithm will be evaluated on a more challenging real-word applications such as medical image super-resolution and hyperspectral imagery in the future work. Furthermore, a more detailed analysis of the proposed algorithm based on different wavelet and filter parameters will also be carried out along with testing the proposed algorithm on other image inverse problems, such as denoising, inpainting, etc.

CONFLICTS OF INTEREST: The authors declare no conflict of interest.

REFERENCES


Awaiz Ahmed received his B.E. degree in Software Engineering from Mehran University of Engineering and Technology, Jamshoro, Pakistan, and M.S. degree in Information and Software Engineering from University of Electronic Science and Technology, Sichuan, China. He is currently working toward the Ph.D. degree in Image Processing, Wavelet processing, Super-resolution, Deep Learning, and Sparse representations.

She Kun received B.Sc. degree from University of Electronic Science and Technology, Sichuan, China. M.S. degree from Southwest Institute of Communications and a Ph.D degree in CS from University of Electronic Science and Technology, Sichuan, China. He is a Professor at University of Electronic Science and Technology of China, Sichuan, Chengdu, China. His research interests include Intelligent Computing, Cloud Computing and Big Data, Network Security and Network Engineering.

Rahael Ahmed Memon Mr. Memon is pursuing his PhD in the school of Computer Science & Engineering at University of Electronic Science and Technology of China. He received his Master’s degree in Computer Engineering from Myongji University South Korea in year 2012. He is also a fulltime faculty member at Computer Science Department Sukkur IBA University, Pakistan. His research interest includes: Internet of Things, image and video processing, Fault tolerant networking, Embedded Systems, and NAND Flash Memories.

Junaid Ahmed received his B.E. degree in Telecommunications Engineering from Mehran University of Engineering and Technology, Jamshoro, Pakistan (2006-2010), and M.S. degree in Electrical and Electronics Engineering from Eastern Mediterranean University(EMU), North Cyprus, Turkey (2014-2015). He is currently working toward the Ph.D. degree in Non-Destructive Testing and Structural Health Monitoring using infrared thermography at UESTC, Chengdu, China. He is also a fulltime faculty member at Electrical Department Sukkur IBA University, Pakistan. His current research interests include wavelet processing, super-resolution, quantitative non-destructive testing and evaluation, sparse representations and low rank matrix factorization, tensor decomposition.
Getenet Tefera received B.Sc. degree from Haramaya University, Harar, Ethiopia and M. Tech degree from Symbiosis International University, Pune, India in Computer Science in 2012 & 2016 respectively. He is currently pursuing the PhD degree in Software Engineering at University of Electronic Science and Technology of China, Sichuan, Chengdu, China. His research interests include Multi-Access Edge Computing, Image Processing, Deep Learning, and Block Chain.