IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 22, 2021, accepted May 1, 2021, date of publication May 5, 2021, date of current version May 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3077611

Accurate Magnetic Resonance Image Super-Resolution Using Deep Networks and Gaussian Filtering in the Stationary Wavelet Domain

GUNNAM SURYANARAYANA[1,2], KARTHIK CHANDRAN[3], (Senior Member, IEEE), OSAMAH IBRAHIM KHALAF[4], YOUSEEF ALOTAIBI[5], ABDULMAJEED ALSUFYANI[6], AND SALEH AHMED ALGHAMDI[7]

[1]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Department of Electronics and Communications, V R Siddhartha Engineering College, Vijayawada 520007, India
[3]Department of Mechatronics, Jyothi Engineering College, Thrissur 679531, India
[4]Al-Nahrain Nanorenewable Energy Research Center, Al-Nahrain University, Baghdad 64074, Iraq
[5]Department of Computer Science, College of Computers and Information Systems, Umm Al-Qura University, Makkah 77207, Saudi Arabia
[6]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia
[7]Department of Information Technology, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia

Corresponding authors: Gunnam Suryanarayana (surya_gunnam@yahoo.co.in) and Karthik Chandran (karthikc@jecc.ac.in)

This work was supported by the Taif University, Taif, Saudi Arabia, through Taif University Researchers Supporting under Project TURSP-2020/115.

ABSTRACT In this correspondence, we present an accurate Magnetic Resonance (MR) image Super-Resolution (SR) method that uses a Very Deep Residual network (VDR-net) in the training phase. By applying 2D Stationary Wavelet Transform (SWT), we decompose each Low Resolution (LR)-High Resolution (HR) example image pair into its low-frequency and high-frequency subbands. These LR-HR subbands are used to train the VDR-net through the input and output channels. The trained parameters are then used to generate residual subbands of a given LR test image. The obtained residuals are added with their LR subbands to produce the SR subbands. Finally, we attempt to maintain the intrinsic structure of images by implementing the Gaussian edge-preservation step on the SR subbands. Our extensive experimental results show that the proposed MR-SR method outperforms the existing methods in terms of four different objective metrics and subjective quality.

INDEX TERMS Deep learning, edge-preservation, MR imaging, residual network, stationary wavelet decomposition, super-resolution.

I. INTRODUCTION

The motive of single-image Super-Resolution (SR) is to generate a High Resolution (HR) image from an input Low Resolution (LR) image. The generated HR image is expected to possess adequate edge information with minimum artifacts. Single-image SR algorithms are currently active in industrial and academic applications as well. Few applications include medical imaging, crime investigation, video surveillance, infrared image processing, and consumer electronics [1]–[5]. In this work, we focus on improving the resolution of LR Magnetic Resonance (MR) images to benefit clinical applications. As a result, we can overcome the limitations of MR imaging, such as a lower signal-to-noise ratio and longer scan time. Research on single-image SR has been classified into three categories, namely interpolation methods, dictionary learning methods, and deep learning methods.

Bilinear and bicubic methods are the widely used traditional interpolation algorithms in practice [6]. The bilinear method considers four closest neighbor pixels, and the bicubic method considers sixteen pixels to compute the unknown pixels. Every missing pixel in the HR grid is obtained using the neighboring pixels based on simple isotropic kernels. As a result, the bilinear and bicubic methods fail to preserve the intrinsic edge structures leading to severe blurring and jaggy artifacts. Jaggies are unwanted high-frequency components

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

that appear along straight lines or curved edges in the reconstructed image. These artifacts appear so widespread in digital display devices and remain an issue in the design of printing devices. To overcome these deficiencies, edge-directed interpolation methods are proposed [7], [8]. These methods attempt to preserve the overall edge structures of images but often produce speckle-noise around edges. Besides, the performance of these methods degrades very rapidly for scale factors above 2. To address this issue, interpolation techniques are combined with wavelet transforms. The Discrete Wavelet Transform (DWT) [9], [10], Stationary Wavelet Transform (SWT) [11], lifting wavelet transform [12], and dual-tree complex wavelet transform [13] are used for wavelet domain interpolation. Recently, a Rational Fractal Interpolation (RFI) model is constructed by Zhang *et al.* [14] and Shao *et al.* [15] for preserving textural details. The RFI function is more accurate and works better than polynomial interpolation kernels [7], [8]. However, the performance of interpolation methods [6]–[8], [14] is inferior to learning-based SR methods.

Learning-based methods use an external LR-HR training dataset to find the missing high-frequency details in an LR image. These methods can be further categorized into two types. The first category is based on dictionary learning, and the second is based on deep learning. Dictionary learning methods solve the SR problem based on statistical analysis or intuitive understanding of many natural images. Neighbor embedding approaches [16], [17] and sparse coding approaches [18]–[24] are the commonly used dictionary-based methods in practice. In neighbor embedding approaches, an input LR image is usually expressed as a weighted linear combination of example patches from an LR dictionary. The same weight combination is used with its HR dictionary to reconstruct the output patch. This is because LR and HR images share similar local geometric structures. However, this approach leads to a quick expansion of dictionaries when the desired training dataset is large.

On the other hand, sparse coding approaches address this issue by enforcing an efficient representation using a sparse linear combination. Yang *et al.* [18] applied this idea to the SR problem by jointly training the LR-HR dictionaries using similar sparse representations. Their algorithm first computes a sparse representation vector from a given input patch and uses the same sparse prior for computing the output patch. Zeyde *et al.* [19] made significant improvements to Yang *et al.*'s model [18] in terms of speed and quality. Other attempts used centralized, nonlocally centralized sparse representations [20], [21], and statistical prediction model without sparse invariance assumption [22]. Timofte *et al.* [23], [24] attempted to reduce the running time of sparse coding approaches by combining sparse learned LR-HR dictionaries with neighbor embedding approaches. However, these methods fail to yield superior SR results compared to the recent deep learning-based methods [25]–[28].

The second category of learning-based SR methods uses deep learning technology by applying Convolutional Neural Networks (CNNs) [25]–[39]. This class of SR methods has gained considerable attention from many researchers in recent times. Dong *et al.* [29], [30] introduced the first deep convolutional network to solve the SR problem. Motivated by this work, several other problems like depth map SR [31] and face hallucination [32] have produced state-of-the-art results. Unlike the traditional learning-based methods [16]–[24], Dong *et al.*'s method [30] directly learns a non-linear mapping between LR and HR spaces. This end-to-end mapping in deep networks trains all the model parameters more efficiently, leading to an accurate inference. Inspired by this, Wang *et al.* [33] developed a network with a set of cascaded sparse coding networks in each mapping layer. However, the sparse coding solver cannot guarantee the optimal mapping accuracy of this approach.

Dong *et al.* [34] modified their network [29], [30] in terms of the number of mapping layers, filter sizes, and feature dimensions. The resultant network provides fast upscaling with improved accuracy. In addition, it operates directly on LR images without the initial bicubic interpolation. In other approaches, generative adversarial networks were used to recover fine textures and edges [35], [36]. Self-exemplars were used in [37] where LR-HR training examples exploit self-similarity to enhance the output SR quality. Cui *et al.* [38] introduced a deep cascade network for gradual upscaling of LR patches after each layer. A deep joint SR model was developed in [39] using a high complex convolutional auto-encoder network.

In this paper, we exploit the idea of residual learning [26] in the stationary wavelet domain and attempt to preserve the intrinsic structure of images. The input LR image is subjected to 2D SWT for decomposing into its low-frequency and high-frequency subbands. We prefer SWT to overcome the shift variance and inferior directionality of DWT. The decomposed LR subbands are fed forward through the trained VDR-nets using the four input channels to produce corresponding residual subbands. These residuals are added to the LR subbands to yield the SR subbands. Finally, we apply the edge-preservation step on the SR subbands using Gaussian operation and then fuse the resulting subbands to generate the output SR image.

In the following, we discuss the deep learning methods in our related work module in Section 2. The network architecture and SR reconstruction process of our proposed algorithm are given in Section 3. Section 4 describes the image datasets, methods, metrics, and implementation details, followed by a discussion on experimental results. Finally, conclusions are drawn in Section 5.

II. RELATED WORK

More recently, Very Deep Residual networks (VDR-nets) [25]–[28] have shown great improvement over the existing deep learning networks with faster convergence rates and accurate SR performance. Unlike traditional networks, the VDR-net does not directly reconstruct HR images. Instead, it emphasizes on the residuals between LR-HR image

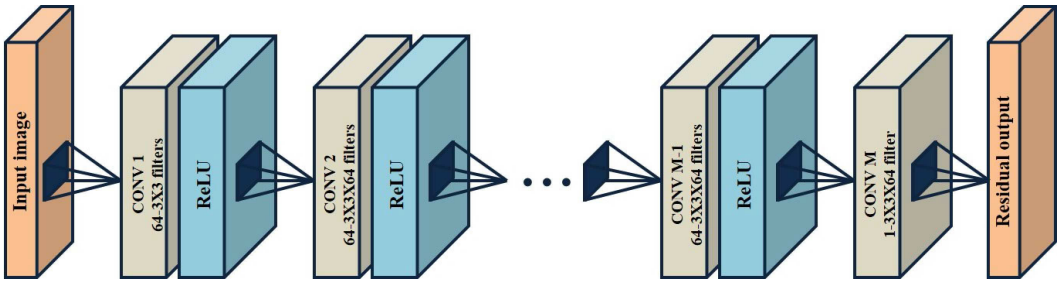

FIGURE 1. VDR-net configuration for prediction of LR-HR output residual image.

pairs and reduces the training time significantly. In particular, we focus on the VDR-net presented by Kim *et al.* [25], [40] in their work.

A. VDR-NET

The structure of the VDR-net is shown in Fig. 1. As illustrated, the network has M layers, namely an image input layer, $M-2$ middle layers, and a regression layer. The filter details of these layers are briefed as follows.

- Image input layer: This is the first layer in the network that operates on input image patches. It consists of 64 filters of size 3×3.
- Middle layers: All the middle layers are similarly parameterized with 64 filters of size $3 \times 3 \times 64$ in each layer.
- Regression layer: This is the last layer in the network which reconstructs the output residual image. It has a single $3 \times 3 \times 64$ sized filter.

All these convolution layers except the regression layer are followed by Rectified Linear Unit (ReLU) layers. The function of ReLU is to introduce nonlinearity in the network by replacing negative values with zeroes. As a result, the time required to estimate the output residual is minimized.

B. DATA MODELING

To compute the output residual for a given LR image, we train the VDR-net using numerous LR-HR example images. Let $\{L^{(i)}, H^{(i)}\}_{i=1}^{N}$ represents the training dataset, where L denotes an interpolated LR image and H represents its HR version. The network learns a model f, which accepts the LR training images and predicts the corresponding HR images. The model can be represented as

$$\hat{H} = f(L), \tag{1}$$

where $\hat{H}$ is an estimate of the original HR image H. The resultant mean square error is modeled as a loss function which has to be minimized by averaging over the training dataset.

$$\text{loss} = \frac{1}{2}\|H - f(L)\|_2^2. \tag{2}$$

In Eq. (2), the mean squared error-l_2 is considered while computing the loss function. l_2 is convex and differentiable in nature which is very advantageous while solving optimization problems. These properties led to l_2's widespread adoption in regression problems, signal and image processing as the dominant error measure.

The objective of the VDR-net is to estimate the residual between input and output images rather than the HR images directly. Now, if $R = H - L$ represents the residual between LR-HR images, the loss function can be modified as

$$\text{loss} = \frac{1}{2}\|R - f(L)\|_2^2. \tag{3}$$

The VDR-net is trained by minimizing the loss function in Eq. (3), and the weights and biases are computed. The Stochastic Gradient Descent with Momentum (SGDM) technique [41] is used for obtaining the optimum parameters of the network. These parameters are used in the testing phase of our method to generate the residual image for a given LR image.

III. PROPOSED METHOD

In this section, we present the details of our network structure, training, and SR reconstruction process.

A. NETWORK STRUCTURE AND LOSS FUNCTION

The task of SR is to restore the maximum lost high-frequency details of an image. An efficient way to achieve this is to process the low-frequency and high-frequency content of an image separately. For this, we apply SWT on a given image I to decompose into its low-frequency LL and high-frequency LH, HL, and HH subbands.

$$\{LL, LH, HL, HH\} = \text{SWT}(I) \tag{4}$$

Eq. (4) represents the subband decomposition of an image I using SWT operation. Here LL represents the approximation coefficient, whereas LH, HL, and HH are the detail coefficients along horizontal, vertical, and diagonal directions. Similarly, for a given training dataset $\{L^{(i)}, H^{(i)}\}_{i=1}^{N}$, the LR subbands are given as $\{LL^L, LH^L, HL^L, HH^L\}$ and the HR subbands as $\{LL^H, LH^H, HL^H, HH^H\}$. Considering all these subbands together, we solve the SR problem in the wavelet domain.

Fig. 2 illustrates the structure of our deep CNN used in the training phase. The network has four input and four output channels connected to the VDR-net. The input channels couple each LR subband with VDR-net, transforming into the corresponding output residual subbands at the output channels. The output residuals corresponding to the

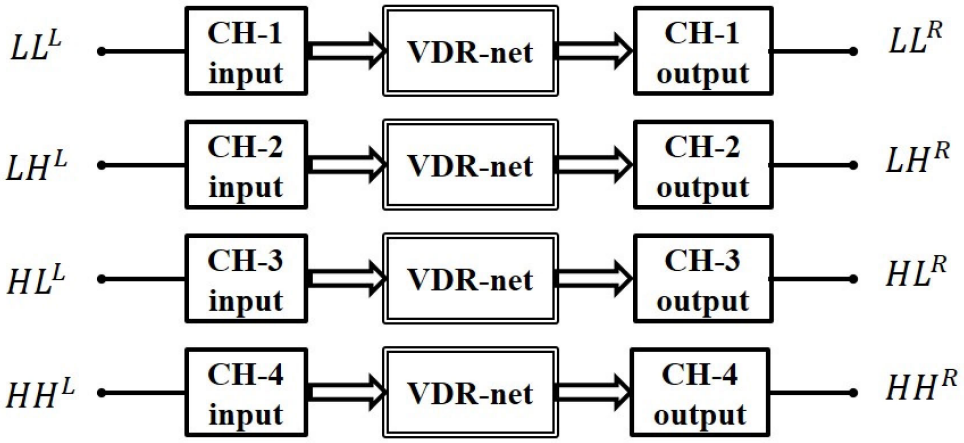

FIGURE 2. Proposed network structure.

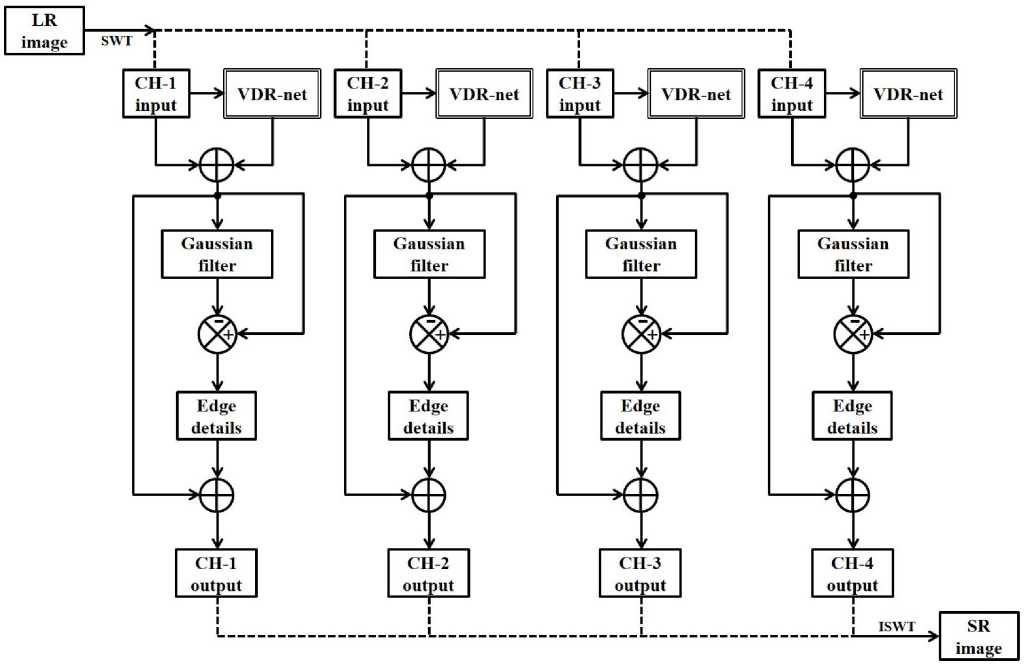

FIGURE 3. Proposed SR reconstruction.

LR subbands $\{LL^L, LH^L, HL^L, HH^L\}$ are represented by $\{LL^R, LH^R, HL^R, HH^R\}$.

As described in Section 2, the VDR-net has an image input layer, $M-2$ middle layers, and a regression layer. The input layer consists of 64 filters of size 3×3, the middle layers have 64 filters of size $3\times3\times64$ in each layer, and the regression layer has a single $3\times3\times64$ sized filter. Here the objective of VDR-net is to train a model f that accepts the LR subbands and outputs the residuals between LR-HR subbands. From Eq. (3), the loss function corresponding to each channel is given as

$$CH_i^{loss} = \frac{1}{2}\|J_i^R - f(J_i^L)\|_2^2, \tag{5}$$

where $J_i \in \{LL, LH, HL, HH\}$ for $i = \{1, 2, 3, 4\}$. We minimize the loss function using the SGDM technique [41] and solve for the optimum network parameters. The momentum and L_2 regularization parameter are set to 0.9 and 0.0001 during the optimization process. The obtained network parameters are used in the testing phase to generate the residual subbands for a given set of LR image subbands.

B. SR RECONSTRUCTION

The proposed SR reconstruction scheme is outlined as a block diagram in Fig. 3. The given LR test image L is first interpolated using the bicubic filter with a factor s and then subjected to 2D SWT decomposition.

From Eq. (4):

$$\{LL^L, LH^L, HL^L, HH^L\} = \text{SWT}(L). \tag{6}$$

The LR subbands $\{LL^L, LH^L, HL^L, HH^L\}$ are fed forward through the VDR-net using the four input channels. The VDR-net transforms the LR subbands into the corresponding

output residual subbands $\{LL^R, LH^R, HL^R, HH^R\}$. By adding these residuals to the LR subbands produces the SR subbands $\{LL^S, LH^S, HL^S, HH^S\}$.

$$J_i^S = J_i^L + J_i^R, \tag{7}$$

where

$$J_i \in \{LL, LH, HL, HH\} \text{ for } i = \{1, 2, 3, 4\}. \tag{8}$$

In the SR problem, bicubic interpolation in the preprocessing stage leads to the loss of high-frequency components. This loss is due to the averaging effect caused by the interpolation filters. As a result, the SR subbands are deficient in edge information causing poor reconstruction quality. To increase the quality of the SR image, preserving edge information is essential. We apply the Gaussian low pass filter on all the SR subbands and subtract the Gaussian smoothed output from the SR subbands.

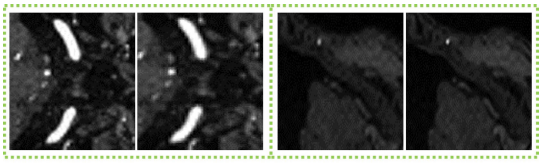

FIGURE 4. SR result pairs on the IXI-MR test images: (a) without Gaussian (b) with Gaussian (c) without Gaussian (d) with Gaussian.

The necessity of Gaussian edge-preservation can be understood from Fig. 4. In Fig. 4(a)-Fig. 4(b), we present one pair of the proposed SR results without and with the Gaussian operation, respectively. Similarly, Fig. 4(c)-Fig. 4(d) represent another pair of our SR results. From these figures, we can see that the SR images in Fig. 4(a) and Fig. 4(c) suffer from poor edge information. Whereas the SR images in Fig. 4(b) and Fig. 4(d) have better edge details and are visually pleasing. The Gaussian process extracts the edge details from each SR subband which are then added back to the SR subbands. The resultant SR subbands have more edge details leading to superior SR reconstruction quality when fused using Inverse SWT (ISWT).

IV. RESULTS AND DISCUSSION

In this section, we conduct experiments to evaluate the performance of our method. First, we present the details about image datasets, methods, and metrics used for comparison. Next, the parameter settings for training and testing are given. Finally, we provide a discussion on the results and assess the quality of SR algorithms.

A. EXPERIMENTAL CONFIGURATION

1) IMAGE DATASETS

The VDR-net is trained using the public IXI-MR image dataset (http://brain-developme-nt.org/ixi-dataset) [42]. The dataset has 600 MR-HR images collected from three different hospitals with Philips $3T$, Philips $1.5T$, and GE $3T$ systems, respectively. We have randomly selected 500 subjects with augmentation for training and performed selective testing on the other 100 subjects.

For a fair comparison with the state-of-the-art methods, we also train the VDR-net by employing three benchmark datasets, namely 91 image dataset from Yang *et al.* [18], BSDS200, and BSDS300 from Berkeley segmentation datasets [43]. A total of 591 HR images are used in the training phase with data augmentation. For SR image reconstruction, we employ two benchmark datasets in the testing phase. 'Set5' [17] and 'Set14' [19] containing 5 and 14 images respectively. Total 19 images are used with three different scaling factors (2, 3, and 4). We use the luminance channel information alone for both training and testing phases. This is because humans are more sensitive to changes in intensity than in color.

2) METHODS AND METRICS

To compare the SR performance, twelve classic and recent state-of-the-art methods are involved in testing. These methods are based on interpolation, dictionary learning and deep learning. Interpolation methods include bicubic, new edge directed interpolation (NEDI) [7], local RFI (LRFI) [15]. SR via sparse representation (SCSR) [18], statistical prediction model based on sparse representations (SPMSR) [22], adjusted anchored neighborhood regression (A+) [24] are dictionary learning techniques. Cascaded deep sparse coding based networks (SCN) [33], SR using deep CNNs (SRCNN) [30], accelerating SRCNN (FSRCNN) [34], SR using very deep CNNs (VDSR) [25], deep wavelet prediction for SR (DWSR) [27], multi-scale saliency and deep CNNs for SR (MSSCNN) [28] are based on deep learning.

To assess the performance of SR methods, four full-reference objective assessment indices are reported. They are Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), perceptual blur (Q_{blur}), and ringing (Q_{ring}) metrics [44]. PSNR and SSIM metrics have been widely in the field of digital imaging for several decades. However, many researchers argue that these simple fidelity metrics do not provide a precise correlation with human perception. Hence, we report two additional metrics Q_{blur} and Q_{ring} along with PSNR and SSIM. Q_{blur} and Q_{ring} metrics enable us to measure the artifacts present in SR images and better assess the algorithms. For better performance, high PSNR, SSIM, low Q_{blur}, and Q_{ring} indices are desired.

B. PARAMETER SETTINGS

1) TRAINING

During the training phase, the images are subjected to one level 2D SWT with Haar wavelet function. It decomposes the HR images into $\{LL^H, LH^H, HL^H, HH^H\}$ subbands. These subbands are then down-sampled and upscaled using the bicubic filter by the same factor s ($s = 2$, 3, and 4) to yield the LR image subbands $\{LL^L, LH^L, HL^L, HH^L\}$. On subtracting the LR subbands from the HR subbands produces the residual subbands $\{LL^R, LH^R, HL^R, HH^R\}$. The VDR-net is trained using the LR and residuals subbands individually for each subband type. The subbands are cropped to 41×41

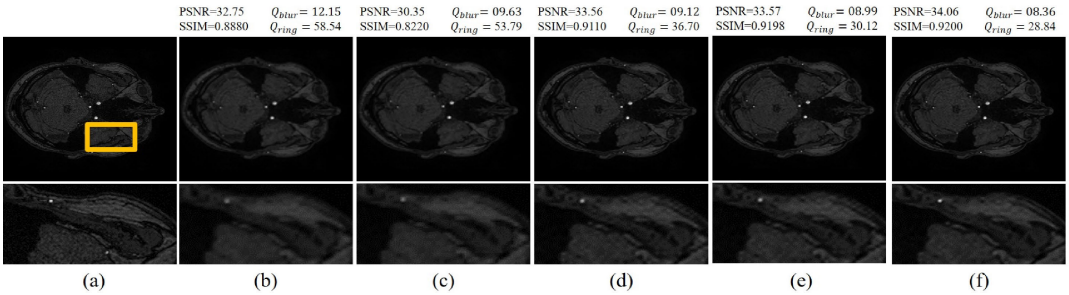

FIGURE 5. Factor 4 SR results on the IXI-MR test image: (a) Ground truth image (b) bicubic (c) LRFI (d) VDSR (e) MSSCNN (f) Proposed.

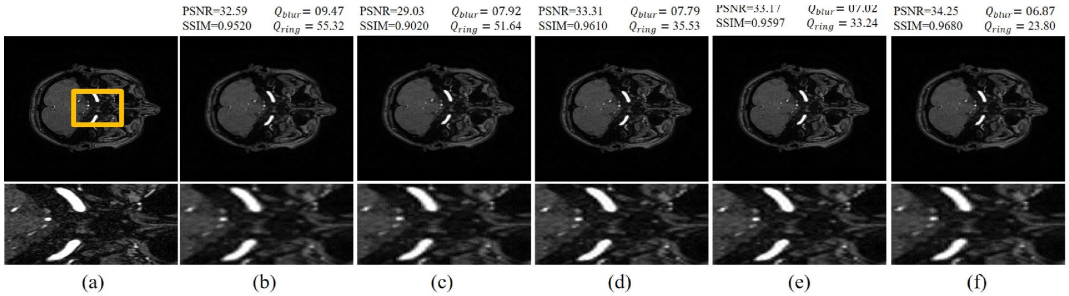

FIGURE 6. Factor 3 SR results on the IXI-MR test image: (a) Ground truth image (b) bicubic (c) LRFI (d) VDSR (e) MSSCNN (f) Proposed.

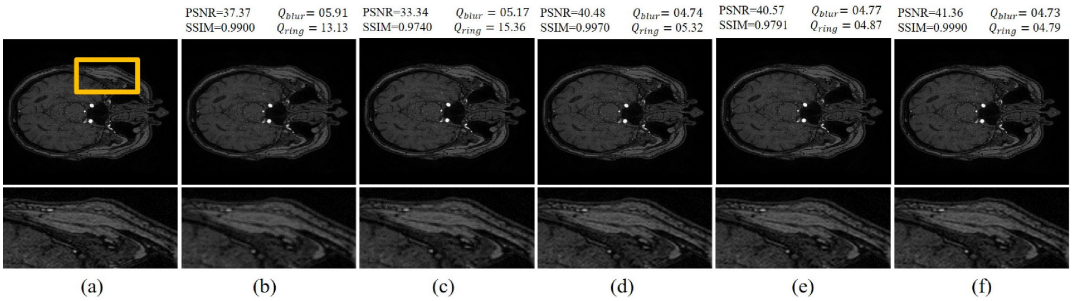

FIGURE 7. Factor 2 SR results on the IXI-MR test image: (a) Ground truth image (b) bicubic (c) LRFI (d) VDSR (e) MSSCNN (f) Proposed.

sub-images with no overlap. The depth of the VDR-net is set to $M = 20$, and the weights and biases are updated by minimizing the loss function in Eq. (5) using the SGDM optimizer. For SGDM optimization, the momentum and L_2 regularization parameter are set to 0.9 and 0.0001, respectively. The gradients threshold is set to 0.01 using norm clipping. The initial learning rate is 0.1 and decreases by a factor 10 for every 10 epochs. After 100 epochs, the VDR-net is fully converged, and it takes about 12 hours to train all the subbands with a single GPU.

2) TESTING

We consider the ground truth images from the IXI-MR dataset, 'Set5' [17], and 'Set14' [19], for full reference image quality assessment. These images are down-sampled and upscaled using bicubic interpolation by a factor s ($s = 2, 3$ and 4) to generate the LR test images. We use one level 2D SWT with the Haar wavelet function for subband decomposition. The Gaussian low pass filtering is applied on the SR subbands with standard deviation $\sigma = 0.9$. For SR reconstruction by a factor s, we use the weights and biases of the VDR-net trained with the same factor s.

C. QUALITY ASSESSMENT

We assess the quality of SR algorithms for three different upscaling factors. Figs. 5-7 present the SR results on three different test images from the IXI-MR dataset for $s = 4$, $s = 3$, and $s = 2$ respectively. Figs. 5(a)-7(a) represent the ground truth MR images. The results based on interpolation are depicted in Figs. 5(b)-7(b) for bicubic and Figs. 5(c)-7(c) for LRFI. In Figs. 5(d)-7(d) and Figs. 5(e)-7(e), we present the results of deep learning methods VDSR and MSSCNN, respectively. When compared with interpolation methods and deep learning methods in Figs. 5(b)-7(b), Figs. 5(c)-7(c), Figs. 5(d)-7(d), and Figs. 5(e)-7(e), the SR images of the proposed method in Figs. 5(f)-7(f) appears close to the ground truth images. Besides, the proposed method has high PSNR and SSIM indices with minimized Q_{blur} and Q_{ring} artifacts.

To prove the effectiveness of the proposed method, we further trained VDR-net using Yang *et al.* [18] dataset, BSDS200, and BSDS300 datasets [43]. The results are tested on standard 'Set5' [17] and 'Set14' [19] images. Figs. 8-13 present visual comparisons of SR methods on 'Set5' and 'Set14' test images. Fig. 8 and Fig. 9 show the SR results for $s = 4$, Fig. 10 and Fig. 11 show the SR results for $s = 3$,

FIGURE 8. Factor 4 SR results on Comic image: (a) Ground truth image (b) bicubic (c) SCSR (d) SPMSR (e) A+ (f) SCN (g) SRCNN (h) FSRCNN (i) VDSR (j) DWSR (k) MSSCNN (l) Proposed.

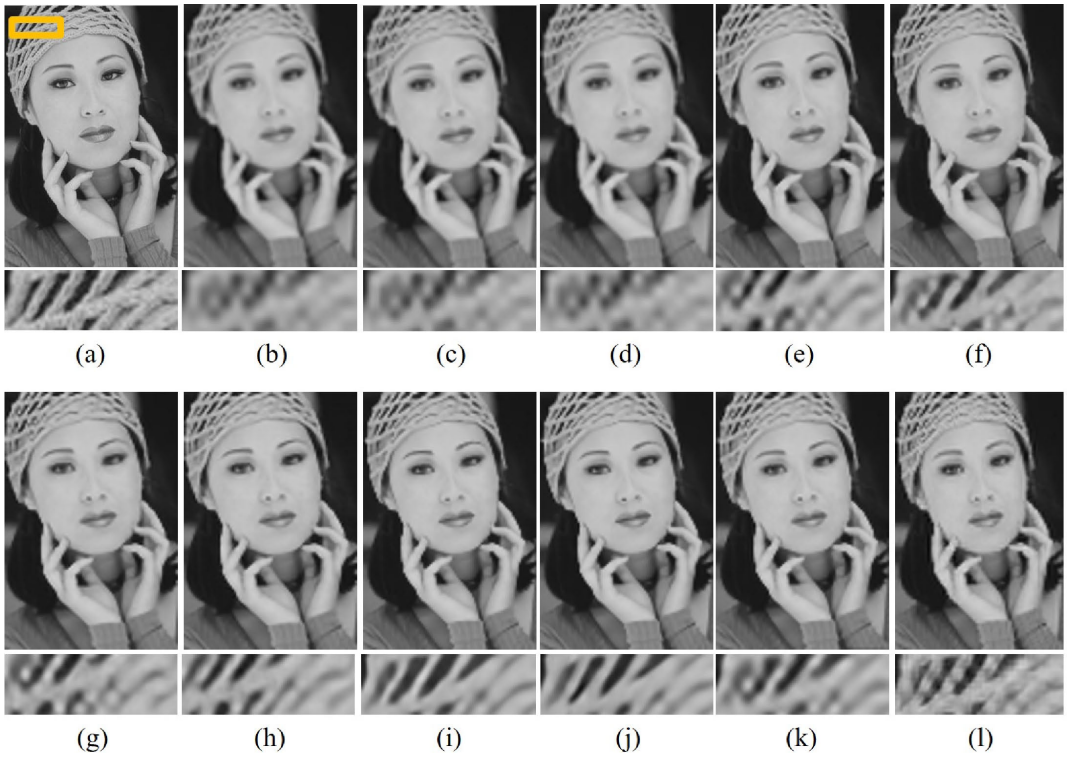

FIGURE 9. Factor 4 SR results on Woman image: (a) Ground truth image (b) bicubic (c) SCSR (d) SPMSR (e) A+ (f) SCN (g) SRCNN (h) FSRCNN (i) VDSR (j) DWSR (k) MSSCNN (l) Proposed.

and Fig. 12 and Fig. 13 show the SR results for $s = 2$. Fig. 8(a) and Fig. 9(a) represent the ground truth images of Comic and Woman images. Fig. 8(b) and Fig. 9(b) are the results obtained using bicubic interpolation. The results of dictionary learning methods, namely SCSR, SPMSR, and A+ are depicted in Figs. 8(c)-8(e) and Figs 9(c)-9(e). Whereas

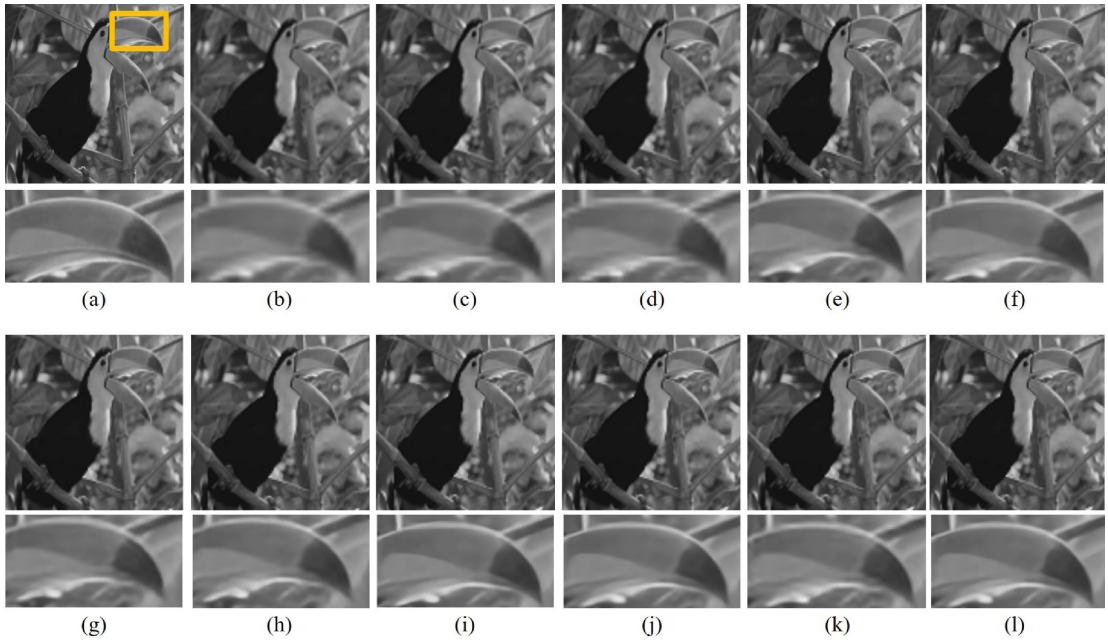

FIGURE 10. Factor 3 SR results on Bird image: (a) Ground truth image (b) bicubic (c) SCSR (d) SPMSR (e) A+ (f) SCN (g) SRCNN (h) FSRCNN (i) VDSR (j) DWSR (k) MSSCNN (l) Proposed.

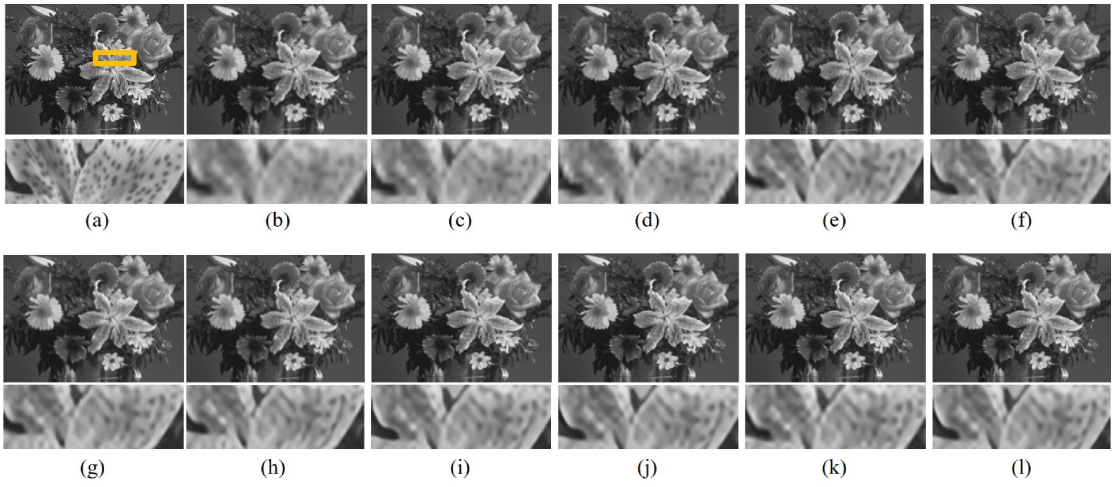

FIGURE 11. Factor 3 SR results on Flowers image: (a) Ground truth image (b) bicubic (c) SCSR (d) SPMSR (e) A+ (f) SCN (g) SRCNN (h) FSRCNN (i) VDSR (j) DWSR (k) MSSCNN (l) Proposed.

the deep learning methods SCN, SRCNN, FSRCNN, VDSR, DWSR, and MSSCNN are shown in Figs. 8(f)-8(k) and Figs. 9(f)-9(k). The SR reconstruction results of the proposed method are outlined in Fig. 8(l) and Fig. 9(l). The bicubic method suffers from blurred and ringing artifacts. The SR images of dictionary learning methods are distorted and fuzzy in nature. The deep learning methods generate images with less distortion and a minimized number of artifacts. However, the details of the flower (Figs. 8(f)-8(k)) and the textures of the hat (Figs. 9(f)-9(k)) in the cropped regions are not well preserved. The flower details of Comic image and the hat textures of Woman image are more efficiently preserved in the proposed method, as shown in Fig. 8(l) and Fig. 9(l).

Figs. 10(a)- 13(a) represent the ground truth images of Bird, Flowers, Baby, and Coastguard images, respectively. The results of different SR methods are shown in Figs. 10(b)- 10(l) and Figs. 11(b)- 11(l) for $s = 3$, Figs. 12(b)- 12(l) and Figs. 13(b)- 13(l) for $s = 2$. Figs. 10(b)-13(b) show the bicubic interpolated images. Figs. 10(c)-13(c), Figs. 10(d)-13(d) and Figs. 10(e)-13(e) represent the dictionary learning methods. The images generated by deep learning methods are given in Figs. 10(f)-13(f), Figs. 10(g)-13(g), Figs. 10(h)-13(h), Figs. 10(i)-13(i), Figs. 10(j)-13(j) and Figs. 10(k)-13(k). The proposed method is outlined in Figs. 10(l)- 13(l). As shown in Figs. 10-13, the bicubic method has severe blurring and ringing artifacts. The performance of dictionary learning methods SCSR, SPMSR, and A+ is inferior to that of the deep learning methods SCN, SRCNN, FSRCNN, VDSR, DWSR, and MSSCNN. However, the edge-preservation in some of these

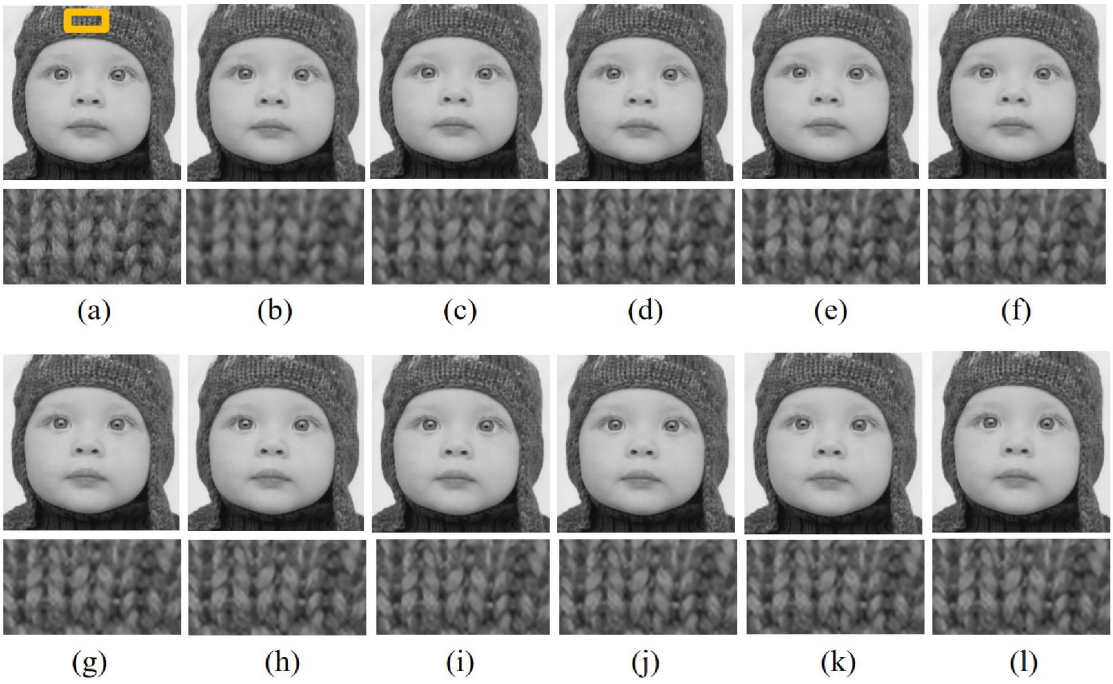

FIGURE 12. Factor 2 SR results on Baby image: (a) Ground truth image (b) bicubic (c) SCSR (d) SPMSR (e) A+ (f) SCN (g) SRCNN (h) FSRCNN (i) VDSR (j) DWSR (k) MSSCNN (l) Proposed.

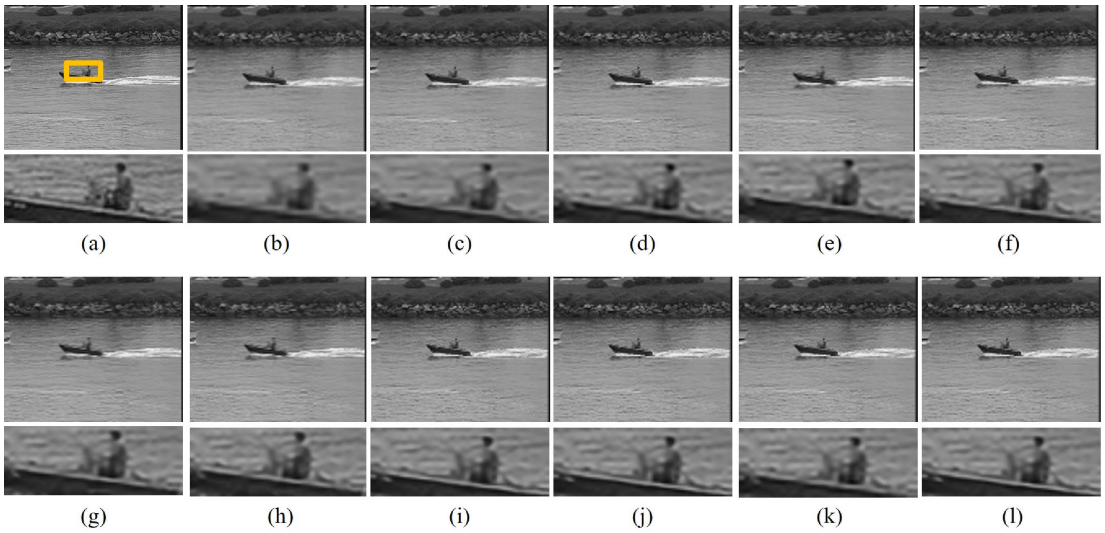

FIGURE 13. Factor 2 SR results on Coastguard image: (a) Ground truth image (b) bicubic (c) SCSR (d) SPMSR (e) A+ (f) SCN (g) SRCNN (h) FSRCNN (i) VDSR (j) DWSR (k) MSSCNN (l) Proposed.

methods is still unsatisfactory (e.g., Fig 10(f), Fig 11(k), and Fig 13(j)). The proposed method has a better performance compared with other methods and is capable of preserving the sharpness of edges (e.g., Fig. 10(l) and Fig11(l)).

Table 1 presents PSNR, SSIM, Q_{blur} and Q_{ring} indices for $s = 4$. Each metric value in the table represents the average of 19 test images from 'Set5' and 'Set14'. Similarly, the average metric indices for $s = 3$ and $s = 2$ are listed in Table 2 and Table 3, respectively.

It can be noticed that the proposed method achieves better results in terms of PSNR for all scaling factors. SSIM index of the proposed for $s = 4$ is slightly inferior to MSSCNN method, but superior for $s = 3$ and $s = 2$. In addition, our method produces low Q_{blur} and Q_{ring} indices for the three scaling factors and is comparable with A+ and deep learning methods. We noticed that LRFI also yields low Q_{blur} values for $s = 2$ and $s = 3$. However, its Q_{ring} values are higher than A+ and the deep learning methods. Also, SCSR and SPMSR methods have low Q_{blur} and Q_{ring} values for $s = 2$, but much higher values for $s = 3$ and $s = 4$. For better performance of an SR algorithm, high PSNR, SSIM, low Q_{blur}, and Q_{ring} indices are desired. Considering all the four objective metrics together, it is evident that the proposed method is quantitatively superior to the existing methods.

In Table 4, we show the running times of different SR methods for $s = 4$. Each value in the table represents the

TABLE 1. PSNR, SSIM, Q_{blur} and Q_{ring} metrics for upscaling factor 4.

Method	PSNR	SSIM	Q_{blur}	Q_{ring}
Bicubic	26.64	0.795	12.94	118.48
NEDI	23.83	0.705	12.31	142.23
LRFI	23.42	0.697	11.02	129.46
SCSR	26.85	0.803	12.34	111.24
SPMSR	26.68	0.797	12.74	117.13
A+	28.09	0.840	09.77	071.73
SCN	28.16	0.839	09.78	076.97
SRCNN	28.29	0.841	09.37	071.65
FSRCNN	28.41	0.844	08.74	061.92
VDSR	28.88	0.855	10.05	078.60
DWSR	28.90	0.856	10.03	079.09
MSSCNN	28.87	0.871	04.96	025.27
Proposed	29.03	0.858	09.79	076.62

TABLE 2. PSNR, SSIM, Q_{blur} and Q_{ring} metrics for upscaling factor 3.

Method	PSNR	SSIM	Q_{blur}	Q_{ring}
Bicubic	28.29	0.864	10.09	74.08
LRFI	24.82	0.766	06.41	49.88
SCSR	28.87	0.878	09.41	62.41
SPMSR	27.97	0.871	09.24	62.70
A+	30.03	0.901	08.02	48.11
SCN	30.04	0.899	08.25	51.26
SRCNN	30.20	0.902	07.65	44.69
FSRCNN	30.41	0.905	07.44	42.18
VDSR	30.79	0.909	07.98	48.06
DWSR	29.99	0.902	08.04	47.79
MSSCNN	30.51	0.907	07.73	45.32
Proposed	30.87	0.910	07.74	45.25

TABLE 3. PSNR, SSIM, Q_{blur} and Q_{ring} metrics for upscaling factor 2.

Method	PSNR	SSIM	Q_{blur}	Q_{ring}
Bicubic	31.14	0.941	7.19	30.77
NEDI	28.39	0.898	7.46	41.25
LRFI	27.75	0.877	5.80	31.97
SCSR	32.60	0.955	6.44	22.69
SPMSR	33.05	0.961	5.96	19.23
A+	33.39	0.964	6.20	19.09
SCN	33.43	0.963	6.16	18.75
SRCNN	33.56	0.964	5.64	16.43
FSRCNN	33.78	0.965	5.67	15.79
VDSR	34.18	0.966	6.06	17.74
DWSR	34.26	0.967	6.07	18.18
MSSCNN	34.29	0.968	5.92	16.35
Proposed	34.57	0.969	5.81	16.84

average of 19 test images. The experiments are conducted in Matlab on a system with 8 GB RAM and Intel(R) Core(TM) i5-7400 CPU: 3.00 GHz. It is noticed that bicubic, A+, DWSR, and MSSCNN are the fastest methods that can execute in 1 second. SCN, FSRCNN, VDSR, and the proposed method consume less than 5 seconds. NEDI, SPMSR, and SRCNN require 10 seconds approximately. LRFI and SCSR are the slowest methods, with execution times 37 seconds and 117 seconds.

TABLE 4. Comparison of running times for upscaling factor 4.

Method	Bicubic	NEDI	LRFI	SCSR	SPMSR	A+
Time	0.22	7.30	36.20	116.96	11.39	0.38
SCN	SRCNN	FSRCNN	VDSR	DWSR	MSSCNN	Proposed
1.70	9.23	2.53	4.96	0.11	0.78	4.22

TABLE 5. Comparison of running times with 1024 × 1024 pixel images for upscaling factor 4.

Method	Bicubic	NEDI	LRFI	SCSR	SPMSR	A+
Time	0.61	22.35	38.20	597.67	32.65	1.46
SCN	SRCNN	FSRCNN	VDSR	DWSR	MSSCNN	Proposed
6.72	25.25	9.36	65.00	0.32	0.99	6.10

In Table 5, we show the running times of different SR methods for $s = 4$. However, all the 19 test images are first resized to 1024 × 1024 before downsampling and upscaling operation by a factor s. We can notice that bicubic, A+, DWSR, and MSSCNN execute in 1 second approximately. SCN, FSRCNN, and the proposed method require less than 10 seconds. NEDI, LRFI, SPMSR, and SRCNN have an execution range between 20 to 50 seconds. Whereas VDSR and SCSR demand 65 seconds and 598 seconds.

From the discussion based on Tables 1-3 and Figs. 5-13, we summarize that our method is superior in terms of objective and subjective quality assessment. Besides, the low computational time is advantageous for real-time implementation.

V. CONCLUSION

In this work, we presented a new single-image MR-SR algorithm using the VDR-net in the stationary wavelet domain. The idea of residual learning and the wavelet subbands increase data sparsity in the training and testing phases. As a result, our algorithm has less computational complexity and hence suitable for 24fps real-time implementation. SWT promises shift-invariance and superior directionality features when compared to DWT. Besides, the edge-preservation using Gaussian operation helps to maintain the intrinsic structure of the SR images. We have shown improvements over the conventional and state-of-the-art SR methods in PSNR, SSIM, Q_{blur}, and Q_{ring} metrics. In addition, the edge-preserving nature of the proposed SR method can be identified from the subjective analysis.

REFERENCES

[1] J. Tian and K.-K. Ma, "A survey on super-resolution imaging," *Signal, Image Video Process.*, vol. 5, no. 3, pp. 329–342, 2011.

[2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, Oct. 2003.

[3] M. Krichen, S. Mechti, R. Alroobaea, E. Said, P. Singh, O. I. Khalaf, and M. Masud, "A formal testing model for operating room control system using Internet of Things," *Comput., Mater. Continua*, vol. 66, no. 3, pp. 2997–3011, 2021.

[4] O. I. Khalaf, K. A. Ogudo, and M. Singh, "A fuzzy-based optimization technique for the energy and spectrum efficiencies trade-off in cognitive radio-enabled 5G network," *Symmetry*, vol. 13, no. 1, p. 47, Dec. 2020.

[5] X. Xiang, Q. Li, S. Khan, and O. I. Khalaf, "Urban water resource management for sustainable environment planning using artificial intelligence techniques," *Environ. Impact Assessment Rev.*, vol. 86, Jan. 2021, Art. no. 106515.
[6] H. Hou and H. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 6, pp. 508–517, Dec. 1978.
[7] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Mar. 2001.
[8] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.
[9] H. Chavez-Roman and V. Ponomaryov, "Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1777–1781, Oct. 2014.
[10] Y. Chen, K. Niu, Z. Zeng, and Y. Pan, "A wavelet based deep learning method for underwater image super resolution reconstruction," *IEEE Access*, vol. 8, pp. 117759–117769, 2020.
[11] G. Suryanarayana and R. Dhuli, "Super-resolution image reconstruction using dual-mode complex diffusion-based shock filter and singular value decomposition," *Circuits, Syst., Signal Process.*, vol. 36, no. 8, pp. 3409–3425, 2017.
[12] M. V. Rao and V. B. Raju, "Image resolution enhancement technique using lifting wavelet and discrete wavelet transforms," in *Innovations in Computer Science and Engineering*. Hyderabad, India: Springer, 2016, pp. 235–239.
[13] M. Z. Iqbal, A. Ghafoor, A. M. Siddiqui, M. M. Riaz, and U. Khalid, "Dual-tree complex wavelet transform and SVD based medical image resolution enhancement," *Signal Process.*, vol. 105, pp. 430–437, Dec. 2014.
[14] Y. Zhang, Q. Fan, F. Bao, Y. Liu, and C. Zhang, "Single-image super-resolution based on rational fractal interpolation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3782–3797, Aug. 2018.
[15] K. Shao, Q. Fan, Y. Zhang, F. Bao, and C. Zhang, "Noisy single image super-resolution based on local fractal feature analysis," *IEEE Access*, vol. 9, pp. 33385–33395, 2021.
[16] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jul. 2004, p. 1.
[17] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. BMVC*, 2012, pp. 1–5.
[18] 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.
[19] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. curves Surf.*, 2010, pp. 711–730.
[20] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1259–1266.
[21] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.
[22] 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.
[23] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.
[24] R. Timofte, V. De, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 111–126.
[25] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Oct. 2016, pp. 770–778.
[27] T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga, "Deep wavelet prediction for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 104–113.
[28] G. Suryanarayana, E. Tu, and J. Yang, "Infrared super-resolution imaging using multi-scale saliency and deep wavelet residuals," *Infr. Phys. Technol.*, vol. 97, pp. 177–186, Mar. 2019.
[29] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.
[30] 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. 2016.
[31] T.-W. Hui, C. C. Loy, and X. Tang, "Depth map super-resolution by deep multi-scale guidance," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 353–369.
[32] S. Zhu, S. Liu, C. C. Loy, and X. Tang, "Deep cascaded bi-network for face hallucination," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 614–630.
[33] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. Int. Conf. Comput. Vis.*, 2015, pp. 370–378.
[34] 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.
[35] C. Ledig, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 4681–4690.
[36] X. Wang, K. Yu, C. Dong, and C. Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 606–615.
[37] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5197–5206.
[38] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen, "Deep network cascade for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 49–64.
[39] Z. Wang, Y. Yang, Z. Wang, S. Chang, W. Han, J. Yang, and T. Huang, "Self-tuned deep super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 1–8.
[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[41] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Oct. 1998.
[42] Y. Huang, L. Shao, and A. F. Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5787–5796.
[43] D. Martin, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. ICCCV*, Vancouver, BC, Canada, 2001, pp. 1–10.
[44] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 163–172, Feb. 2004.

GUNNAM SURYANARAYANA received the B.Tech. and M.Tech. degrees from Jawaharlal Nehru Technological University, India, in 2008 and 2010, respectively, and the Ph.D. degree from the School of Electronics Engineering, VIT University, Vellore, India, in 2016. He is currently a Postdoctoral Researcher with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China. His research interests include image super-resolution, image fusion, and medical imaging.

KARTHIK CHANDRAN (Senior Member, IEEE) received the B.E. degree in electronics and instrumentation engineering from the Kamaraj College of Engineering and Technology, India, in 2007, and the M.Tech. degree in control and instrumentation engineering and the Ph.D. degree from the Kalasalingam Academy of Research and Education (KARE), in 2011 and 2017, respectively. He served as a Postdoctoral Researcher with Shanghai Jiao Tong University, China, from 2018 to 2020. He is currently serving as an Associate Professor with the Department of Mechatronics Engineering, Jyothi Engineering College, Thrissur. He has published several international journals and conference papers. His research interests include time delay control problem, nonlinear system identification, cascade control systems, and unmanned vehicle.

OSAMAH IBRAHIM KHALAF received the B.Sc. degree in software engineering from Al-Rafidain University College, Iraq, the M.Sc. degree in computer engineering from Belarusian National Technical University, and the Ph.D. degree in computer networks from the Faculty of Computer Systems and Software Engineering, University Malaysia Pahang. He is currently a Senior Engineering and a Telecommunications Lecturer with Al-Nahrain University. He has hold 17 years of university-level teaching experience in computer science and network technology, and has a strong CV about research activities in computer science and information technology projects. He has had many published articles indexed in ISI/Thomson Reuters and has also participated and presented at numerous international conferences. He has a patent and has received several medals and awards due to his innovative work and research activities. He has good skills in software engineering, including experience with Net, SQL development, database management, mobile applications design, mobile techniques, Java development, android development, IOS mobile development, cloud system and computations, and Website design. He was the editor in chief and a main guest editor in many Scopus and SCI index journals. His brilliant personal strengths are in highly self-motivated team player who can work independently with minimum supervision, strong leadership skills, and outgoing personality. He has overseas work experiences in University with Binary University, Malaysia, and University Malaysia Pahang.

YOUSEEF ALOTAIBI received the master's degree in information technology (computer network) from La Trobe University, Melbourne, Australia, in 2009, and the Ph.D. degree from the Department of Computer Science and Computer Engineering, La Trobe University, in 2014. He is currently an Associate Professor with the Department of Computer Science, College of Computer and Information Systems, Umm Al-Qura University, Saudi Arabia. He has published several international journals and conference papers. His research interests include business process modelling, business process reengineering, information systems, security, business, and IT alignment, software engineering, system analysis and design, sustainability, and smart cities development.

ABDULMAJEED ALSUFYANI received the bachelor's degree (Hons.) in computer science from Taif University, Saudi Arabia, in 2006, and the master's and Ph.D. degrees in computer science from the University of Kent, U.K., in 2010 and 2015, respectively. He is currently an Associate Professor of computer science with the College of Computers and Information Technology, Taif University. His research interests include computational intelligence, computational neuroscience, and machine learning algorithms.

SALEH AHMED ALGHAMDI received the Bachelor of Education degree (Hons.) from the Department of Computer Science, Teachers College, Riyadh, Saudi Arabia, in 2004, the Master of Information Technology degree from La Trobe University, Melbourne, Australia, in 2010, and the Doctor of Philosophy degree in computer science from the Royal Melbourne Institute of Technology (RMIT) University, Melbourne, in 2014, thesis title A Context-aware Navigational Autonomy Aid for the Blind. He is currently an Associate Professor with the Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia. His research interests include context awareness, positioning and navigation, and visually impaired assistance.