On learning based compressed sensing for high resolution image reconstruction

Sheikh Rafiul Islam1 | Santi P. Maity1 | Ajoy Kumar Ray2

1 Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal, India
2 Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, West Bengal, India

Correspondence
Santi P. Maity, Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah-711 103, West Bengal, India.
Email: santipmaity@it.iiests.ac.in

Abstract
Compressed sensing (CS) or compressive sampling has shown an enormous potential to reconstruct a signal from its highly under-sampled observations. A high dimensional image processing system can adopt the CS paradigm to reduce the storage and the transmission burden. However, a large sensing system (matrix) is required to capture high dimensional images. The CS reconstruction algorithms demand heavy computational requirement due to the use of large sensing matrix. It sometimes becomes impractical to implement a sensing system of the desired size due to the limited access to the CPU memory. To address this issue, the present work proposes a deep learning based CS framework that uses a convolutional neural network to enable capturing of a high dimensional image by utilizing a multi-layer subsampling and filtering operations. The proposed approach uses another convolutional neural network that reconstructs the original image without depending on the sensing network at a significantly reduced computational cost. Extensive simulation results show that the proposed method reconstructs a high dimensional image with an improved PSNR value by $1.5 \pm 0.63$ dB, SSIM value by $0.04 \pm 0.02$ and FSIM value by $0.02 \pm 0.02$ compared to the other state-of-the-art methods in less than 1.4 seconds.

1 | INTRODUCTION

Compressive sensing (CS) enables reconstruction of a signal from its far fewer number of samples (or measurements) than required by the conventional Shannon–Nyquist sampling theorem [1, 2]. The CS method offers a significant reduction in signal acquisition time, transmission burden and storage requirement. Such signal processing approach can be used in high dimensional imaging applications like optical coherence tomography for retina [3], high resolution radar [4], infrared imaging [5] etc. However, CS requires a very large sensing system or matrix for capturing the high dimensional images. To recover the original signal, the CS reconstruction techniques need solving of an ill-posed inverse problem through the optimizations technique that demand high computation time. The computation time increases proportionally with the size of the sensing matrix. A deep learning (DL) approach can be used efficiently in the CS framework to capture the high resolution images followed by artifacts-free reconstruction at relatively low computation cost. The DL is a multi-layer artificial neural network (ANN) having the capability of extracting the features from the raw input data through the multiple levels of hierarchical abstractions [6, 7].

Conventional CS methods exploit the spatio-temporal correlations to recover the desired image from the highly undersampled observations. To ensure successful recovery of the original signal from its undersampled measurements, the CS theory requires the sensed signal to be sparse, i.e., signal consists of very few nonzero entries, in some representation space using the proper basis. The CS theory also requires that the sensing system to be incoherent i.e. to have very low correlations, with the given basis function [1, 2]. A tractable convex optimization technique enables the recovery of the signal if the sensing system satisfies the restricted isometry property (RIP) [8]. The RIP provides sufficient conditions to ensure the required incoherence between the sensing system and the sparsifying basis, which is mandatory for a faithful reconstruction of the signal from the highly undersampled measurements. A random Gaussian matrix, as used in [1], is usually chosen as sensing system which satisfies the RIP with a minimum number of measurements [8]. Zhang et al. [9] address the CS recovery problem...
by enforcing sparsity and non-local self-similarity simultaneously as a group sparse representation (GSR) for an image. A patch matching based multitemporal (MT) GSR [10] is done in the optical remote-sensing for CS reconstruction of the missing information in the remote sensing images. The method makes use of both the local and the nonlocal correlations effectively in the reconstruction process. Shi et al. [11] apply the GSR in low lighting imaging to improve its quality through the refining of transmission map by a guided image filtering. However, the GSR based approaches exploit the concept of group sparse representation, which is a time consuming reconstruction process.

Li et al. [12] propose a MT dictionary learning based CS that was extended on k-means algorithm and singular value decomposition for the recovery of the quantitative data contaminated by thick clouds and shadows. Shen et al. [13] propose an algorithm that retrieves the dead pixel stripes by adaptive spectrum weighted sparse Bayesian dictionary learning technique to recover aqua moderate resolution imaging spectroradiometer band 6. The method efficiently uses the beta process factor analysis to find the latent spectral correlations among the different spectral bands. However, the dictionary learning based methods are computationally expensive as a large number of dictionary elements of typical interest need to be explored. Eslahi et al. [14] report an adaptive curvelet thresholding criterion for the removal of perturbation during reconstruction. The work also exploits a new sparsity measure called joint adaptive sparsity regularization that enforces the local sparsity and the nonlocal 3D sparsity in transformed domain simultaneously. An efficient alternating minimization technique is used for recovering a highly undersampled signal, however, the regularization-based minimization technique is still time consuming. An edge preserving regularization for total variation (TV) based CS reconstruction is reported in [15] that recovers the low-contrast images from the highly undersampled data. The reconstruction time can be reduced by enhancing the sparsity of the signals using iteratively re-weighted minimization techniques as described in [16, 17]. However, in practice, it is often infeasible to accomplish a large random sensing matrix to capture a high dimensional image. To address the issue, a Block based CS (BCS) is utilized in [18, 19], where the high dimensional image is divided into several non-overlapping blocks and a small random matrix is used repeatedly as sensing system. However, in the BCS approach, the reconstruction often suffers from the blocking artifacts and remains present even after performing some post-processing operations.

In the recent years, DL is used extensively in single image super-resolution problem. Such DL frameworks [6, 7] enables one pass CS reconstruction through the deployment of a trained ANN instead of using the time consuming iterative conventional CS methods. Dong et al. [20] propose a deep convolutional neural network (CNN) that learns mapping of a low-resolution input image to a high-resolution output image. The method uses a deep convolutional network for a sparse-coding based super-resolution imaging to achieve the faster reconstruction speed. Quan et al. [21] propose a novel approach using RefineGAN, a variant of fully-residual convolutional autoencoder and generative adversarial network (GAN) [22] towards achieving a high dimensional magnetic resonance image (MRI) reconstruction in a very short time. The approach outperforms the state-of-the art methods in terms of reconstruction time and quality. Sun et al. [23] use a deep information sharing network in CS-MRI inversion problems, consisting of a densely cascaded inference block, each one containing a feature sharing unit and a data fidelity unit. The network achieves significant improvement in terms of computational speed as well as reconstruction quality. In [24], a self-supervised learning strategy is used, that produces high-resolution dynamic MRI from the sub-sampled data. The method defines loss function in terms of validation subset while the training subset is used to enforce data consistency. The proposed method performs similar to the conventional supervised learning approaches.

Lee et al. [25] use a residual learning approach through deep CNN to accelerate MRI reconstruction from the sub-sampled observations. The method learns to provide an artifacts-free MRI without requiring heavy computation as seen in conventional CS reconstruction. However, the method considers the images suffered from aliasing artifacts obtained due to the direct inversion of the under-sampled observations as the input to the networks. In [26], a residual learning method with a recursion structure is introduced that yields a high resolution image at reduced computational cost. The method optimizes the number of recursions in the feature extraction process using a simplified network structure. Shi et al. [27] propose a CS framework using CNN that overcomes the problem of the random sampling matrix required in the conventional CS reconstruction method. However, the sampling network used in the CS framework is trained to learn a set of pre-defined matrices which may fail to provide a generalized solution. In [28, 29], DL based convolutional CS methods are used to optimize both the sensing and the reconstruction networks together through the end-to-end training. The methods reconstruct a high resolution images at low computation time. In [30], a deep residual learning approach is adopted to speed up the training process while preserving high-frequency details for ensuring the desired reconstruction quality. However, these methods use a single feed forward network for representing the latent features map as well as reconstructing the desired image. The use of such networks may lead to a limited scope if the measurements are observed independently.

Many high dimensional CS imaging applications like X-ray tomography, ultrasound imaging, radar imaging etc. involve high computational time to achieve the desired resolution on reconstructed images. In conventional CS, a large sensing matrix is needed to capture the high dimensional images. This leads to increased CS reconstruction time due to the involvement of large sensing matrix. Reconstruction time increases further if the input measurements are noisy. Apart from the practical feasibility issue of implementing a large sensing system, the main challenge here is to reduce the reconstruction time. To reduce the reconstruction time, a paired DL network may be used. The DL framework optimizes both the linear sensing matrix and the non-linear reconstruction method through an end-to-end training [31]. A trained DL network can capture the high
resolution images efficiently like the conventional CS does whereas the reconstruction is done at a much faster speed compared to the conventional one [20]. However, most of the existing DL based CS frameworks are either block based or use a single network for both the sensing and the reconstruction purposes. In practice, similar to the BCS, the use of a single network is also not preferable due to the inability of using the sensing network independently. Furthermore, the use of same sensing matrix may be a communication overhead in case of far end CS reconstruction. Hence, it is essential to have a CS framework that enables capturing of a high dimensional image and consequently reconstruction of the original image at a negligible computation cost.

To address the above mentioned issues, this paper proposes a GAN inspired DL framework to capture high resolution images using very less number of measurements. This enables artifacts-free reconstruction at low computation cost compared to the conventional CS. The DL approach uses a progressive CNN architecture that helps to capture a high dimensional image with a low dimensional vector through a number of convolutional and pooling layers. The architecture is able to handle spatial domain inputs (i.e. no transformation for sparse data as required in conventional CS) of any size, without using any type of random or other sensing matrices. Another CNN architecture is also used to reconstruct the original image from the low dimensional features vector through a number of convolutional and unpooling (upsampling) layers. Required training for the DL architectures is provided in an adversarial manner as described in [22]. A dataset containing the same types of images that need to be processed by the DL network is used in the training. The overall contributions of this work are summarized as follows: (i) High dimensional images can be captured through a compressed features map or measurements vector without fulfilling any size constraint on the sensing system. (ii) Successful reconstruction is made possible even if the capturing signal doesn’t meet certain sparsity level as required in the conventional CS i.e. the need of satisfying RIP requirement can be avoided. (iii) Extensive simulation results show an improved reconstruction quality in peak signal-to-noise ratio (PSNR), mean structural similarity (SSIM) index and feature similarity index (FSIM) by 1.5 ± 0.63 dB, 0.04 ± 0.02 and 0.02 ± 0.02, respectively over the state-of-the-art methods [9, 19, 20, 26].

The rest of the paper is organized as follows: a detailed description of the proposed approach with network architectures are presented in Section 2. Simulation results and discussion are made in Section 3. Finally, conclusions and the scope of the future work are stated in Section 4.

2. PROPOSED LEARNING BASED CS SENSING AND RECONSTRUCTION

This section presents the proposed learning based sensing and reconstruction approach with a schematic block diagram shown in Figure 1. The sensing network is used to provide the compressed features (as a measurement vector) map of any high dimensional input signal. A trained reconstruction network provides the desired high dimensional image from this measurement vector. The training of the reconstruction network is done in presence of a discriminator network as shown in the Figure 1. The reconstruction network is trained to minimize the reconstruction error that in other way minimizes the probability of assigning the correct label by the discriminator network. The discriminator network is trained to maximize the probability of assigning the correct label to both the reconstructed and the original image.

2.1 Sensing system to obtain CS measurements

In CS, a sparse signal $x \in \mathbb{R}^N$ can be represented by a measurement vector $y \in \mathbb{R}^M$ (where $M << N$) using the following equation [1]

$$y = \Phi x$$

(1)

where $\Phi \in \mathbb{R}^{M \times N}$, represents a sensing matrix.
With the consideration of additive noise presence in the sensing system, the above equation is rewritten as follows:

\[ y = \Phi x + \eta \]  

(2)

where \( \eta \in \mathbb{R}^M \).

Now, the sensing matrix can equivalently be represented using a set of multilayer convolutional filter followed by sub-sampling operator as stated below.

\[
y = S^L \left( f^L(\phi^1, \phi^2, \phi^3, \ldots) \bigotimes S^{L-1} \left( f^{L-1}(\phi^{L-1}_1, \phi^{L-1}_2, \phi^{L-1}_3, \ldots) \right) \right) + \eta
\]

Here \( f^l(\phi^1, \phi^2, \phi^3, \ldots) \) represents the \( l \)-th-layer filter with optimally trained set of kernels \( \phi^1, \phi^2, \phi^3, \ldots \) of size \( k_W \times k_H \). The \( l \)-th-layer sub-sampling operator \( S^l \) of size \( p_W \times p_H \) is used to reduce the spatial dimension of the features for the next layer.

Equation 3 can be easily implemented by a multi-layer convolutional network accompanied by a number of sub-sampling layers. The output features map provided by the final layer i.e. \( L \)-th-layer equivalently represents the CS measurement vector \( y \in \mathbb{R}^M \). The convolutional and the sub-sampling layers essentially maps a high dimensional input signal to a low dimensional CS measurements feature space.

2.1.1 Sensing network architecture

The sensing network uses a number of convolutional layers followed by fully-connected (FC) or dense layers as described in the Table 1. The main objective of the sensing network is to provide highly compressed features map of the input image to be captured. A pooling layer serves the objective through reducing the spatial dimension of the input, while a convolutional layer extracts features from the input. In the proposed sensing network, every convolutional layer (except the final layer) is followed by a pooling layer with a window of size \( 2 \times 2 \) to reduce the input spatial dimension each time by 50%. However, the output of a convolutional layer is kept unchanged by utilizing zero padding with the input. A flatten layer is used to convert multidimensional input features map into a one dimensional vector to be considered as the CS measurements vector. The network uses leaky rectified linear unit (LReLU) and Linear activation functions for the hidden and the final layers, respectively.

The sensing network is trained to provide the best possible measurement vector as features map at high reduction factors. The network considers an input image of size \( 128 \times 128 \times 3 \) and provides a measurements vector of size 512 which is reduced by a factor of 0.01 (approximately) with respect to the full-scale image. However, the network can also be used for other reduction factors by changing the configuration of the final layer. The configuration of the 1st layer also needs to be changed if the dimension of the input image changes. There is no particular guidelines to choose the number of convolutional and the pooling layers for the networks. However, depending upon the dimension of the input image and the reduction factor, this work chooses the number of the layers that tries to reduce the spatial dimension as well as extract features gradually.

2.2 Reconstruction system for CS recovery

The present system uses a trained convolutional network to reconstruct the original image from the CS measurement vector \( y \in \mathbb{R}^M \). The network performs a number of deconvolution and up-sampling operations to yield the desired high dimensional image from the low dimensional CS measurements. This reconstruction network can be designed independently i.e. the number of layers and their types can be defined without following the reverse order of the sensing network.

| Table 1: Sensing network summary |
|----------------------------------|
| Layer information | Activation function | Output shape | No. of parameters |
|-------------------|---------------------|--------------|------------------|
| Input image       | –                   | 128 x 128 x 3| –                |
| Conv (kernel:3 x 3) | LReLU               | 128 x 128 x 16 | 448             |
| Pooling           | –                   | 64 x 64 x 16 | –                |
| Conv (kernel:5 x 5) | LReLU               | 64 x 64 x 32 | 12 832          |
| Pooling           | –                   | 32 x 32 x 32 | –                |
| Conv (kernel:5 x 5) | LReLU               | 32 x 32 x 64 | 51 264          |
| Pooling           | –                   | 16 x 16 x 64 | –                |
| Conv (kernel:3 x 3) | LReLU               | 16 x 16 x 128 | 73 856          |
| Pooling           | –                   | 8 x 8 x 128  | –                |
| Conv (kernel:3 x 3) | LReLU               | 8 x 8 x 128  | 147 584         |
| Flatten (reshape) | –                   | 8129         | –                |
| Dense             | Linear              | 512          | 4 194 816       |
| Total params:     |                     | 4 480 800    |                  |
Training of the reconstruction network is performed in a competitive manner with a discriminator network as described in GAN [22]. The objective of the discriminator network is to maximize the classification accuracy by identifying the original image and the fake image i.e. generated by the reconstruction network correctly. To formulate the objective function of the discriminator network the binary cross-entropy loss function is used as defined below,

\[ \mathcal{L}(\alpha, \beta) = \min_{\alpha} \max_{\beta} \left\{ E_{\zeta \sim p(\zeta)} \left[ \log \beta(\zeta) \right] + E_{\hat{\zeta} \sim p(\hat{\zeta})} \left[ \log (1 - \beta(\hat{\zeta})) \right] \right\} \]  \tag{10} \]

where \( \zeta \) and \( \hat{\zeta} \) are the sample drawn from the probability distribution of the original data \( p(x) \) and the probability distribution of the data obtained from measurement vector \( p(y) \), respectively.

The objective function written in Equation 10 is used as a min-max game where \( \alpha \) tries to minimize the function value and \( \beta \) tries to maximize the value as described [22].

The first term in \( \mathcal{L}(\alpha, \beta) \) is the entropy of the data drawn from the real distribution that is passed through the discriminator. The second term is the entropy of identifying a fake sample by the discriminator. The fake samples are generated by the reconstruction network using the data obtained that is drawn from the probability distribution of \( \zeta \) added with noise.

Learnable parameters of both the networks are updated through gradient descent in an alternating fashion. In the first step, parameters of the discriminator network are updated while the reconstruction network’s parameters are kept unchanged. In the second step, keeping the parameters of the discriminator network unchanged, the reconstruction network is trained to produce images closer to the actual one. Gradients, used in backpropagation for updating the network parameters, are summarized as below [22]:

\[ \nabla_{\theta_{\beta}} \frac{1}{m} \sum_{i=1}^{m} \left[ \log \beta(\zeta^{(i)}) + \log (1 - \beta(\alpha(\zeta^{(i)}))) \right] \]
\[ \nabla_{\theta_{\alpha}} \frac{1}{m} \sum_{i=1}^{m} \left[ \log (1 - \beta(\alpha(\zeta^{(i)}))) \right] \]  \tag{11} \]

where \( \theta_{\beta} \) and \( \theta_{\alpha} \) are the parameters of the discriminator network and the parameters of the reconstruction network, respectively; \( \zeta^{(i)} = \{ \zeta^{(1)}, ..., \zeta^{(m)} \} \) and \( \zeta^{(i)} = \{ \zeta^{(1)}, ..., \zeta^{(m)} \} \) are the \( m \) number of minibatch samples drawn from the distribution of original data points \( p_{\text{data}}(\zeta) \) and the noisy features map provided by the sensing network \( p_{\gamma}(\zeta) \), respectively. Here, \( \zeta \) and \( \hat{\zeta} \) represent the data points drawn from the distribution of the original and the reconstructed images, respectively.
reconstruction of other size images, the final layer’s number of neurons need to be adjusted.

### 2.3 Networks training process

In the proposed approach, the required training of the sensing network and the reconstruction network are provided in presence of the discriminator network as described in GAN [22]. During the training of the reconstruction network, the discriminator network is set as non-learnable (freeze) state. Similarly, during the training of the discriminator network, the reconstruction network is set as non-learnable state. The overall training process is presented as a schematic block diagram shown in Figure 2. The sensing network can be trained along with the reconstruction network as shown in the schematic diagram. However, the sensing network can also be trained completely and independently. In the latter case, a separate dataset containing the latent compressed features vector, provided by a pre-trained (using a small sub-dataset) sensing network, needs to be prepared. During the training process, the output features map provided by the sensing network (or the prepared dataset) is fed to the reconstruction network with some randomly generated additive noise.

The training process of the reconstruction and the discriminator networks is also summarized in Algorithm 1. Both these networks learn based on their previous predictions, competing each other yielding a better outcome. Learnable parameters of the networks are updated through backpropagation of the gradients of the error. This approach uses Stochastic Gradient Descent (SGD) as the optimization method for minimizing the loss function. During the training, the discriminator tries to detect adversaries by updating its parameters constantly. Therefore, the reconstruction network is less likely to be overfitted. A stopping criteria is set by assigning a certain value to the \( N_i \), the total number of iterations.
3 | SIMULATION RESULTS AND DISCUSSION

This section presents the description of the dataset used, system specifications, required tools and assessment results to evaluate performance of the proposed method.

3.1 | Datasets used

An exclusive database is prepared by comprising the real-life images obtained from the retinal image databases DRIVE (digital retinal images for vessel extraction) [32] and STARE (structured analysis of the retina) [33], brain IXI MRI database [34] and chest X-ray image dataset [35] to train the proposed model. The final database is constructed by comprising 2800, 400 and 100 images chosen randomly from the different databases for the training, validation and testing purposes, respectively. All the retina and X-ray images are converted into joint photographic experts group format from portable pixel map and digital imaging and communications in medicine file format, respectively.

The images are resized into \((128 \times 128 \times 3)\) i.e. the desired input size by the network to carry this experiment. However, higher dimension images can also be used for the dataset with a redesigned network architecture and additional expense in training time. All the gray scale (i.e. 1- channel) images are changed into the RGB (i.e. 3- channel) images by repeating the same values for all the channels for simplicity, however, other model can also be used.

3.2 | Experimental setup

The proposed deep NN model is trained in Google’s colabatory with “runtime type” and ‘hardware accelerator’ are chosen as ‘Python 3’ and “graphics processing unit,” respectively. The KERAS 2.0.8 libraries and packages with TensorFlow 1.0 as back-end are used to define the NN model in the colabatory. The results presented for comparative study are performed with MATLAB R2018b, run on a desktop consisted of an Intel Core i5 – 7500 CPU (3.96 GHz turbo), 8GB RAM, windows 10, 64 bit operating system.

3.3 | Training progress

In this experiment, all the learnable parameters i.e. the weights and the biases of the network layers are initialized with the random values chosen over a Gaussian distribution having zero (0) mean and standard deviation 0.001. This work also adopts dropout [36] layer with a probability of 0.5 after the 1st FC layer to help in preventing the data overfitting. The values of the other hyper parameters, used during the training of all the networks, are shown in Table 4. However, these hyper parameters won’t be required once the training of the networks is done, i.e. during the actual performance the hyper parameters aren’t required. The spatial domain images to be sensed is considered as the input to...
the trained sensing network to provide the features vector. The features vector is used as input to the trained reconstruction network to produce the desired image. This work uses medical images to demonstrate the efficiency of the proposed method. However, as the model is trained with the converted format of RGB medical images, the framework can also be used for natural imaging systems if trained with a dataset containing the natural images.

Figure 3 shows the performance of the reconstruction network in terms of the loss evaluated using Equation 8 for both the training and the validation data when reduction factor used is 0.01 for a desired output image of size 128 × 128 × 3. The reconstruction loss goes on reducing with the development of the learning process for both the training and the validation data as expected. The CS measurements obtained by the sensing network are provided as input to the reconstruction network. In the training process, the parameters of both the sensing as well as reconstruction networks get learned. The reduction in loss i.e. difference between the original image and the reconstructed image, with the development of the learning process, confirms the desired performance achievement by the DL network. The training of the networks is provided to reconstruct an image of size 128 × 128 × 3 from a measurement vector of size 512 obtained by considering the reduction factor of 0.01. The different reduction factors are achieved by redefining the number of neurons of the input layer as required. A fine tuning is performed to adapt the changes in the input layer by updating the network parameters retrained over 10 number of epochs.

### 3.4 Performance comparisons

Performance of the proposed method is evaluated in terms of the objective as well as the subjective measures and are compared with other state-of-the-art methods. Both the conventional CS based approaches like GSR [9], BCS [19] and DL based approaches like DeepCNN [26] and ConvNet [20] are considered for the performance comparisons. The widely used PSNR, SSIM index [37] and FSIM [38] metrics are used as quantitative measures to compare the quality of the reconstructed images with respect to the original images. In this experiment, the BCS method [19] considers the size of the block as 32 × 32.

Figures 4 and 5 present some of the reconstructed retina images using the proposed, conventional CS [9, 19] and DL based [20, 26] methods to compare their perceptual quality. Figures 4(a–c) are the reconstructed color retina images using the conventional CS based on GSR [9], BCS [19] and the proposed methods, respectively. Figure 5(a–c) represent the reconstructed retina images using the DL based DeepCNN [26], ConvNet [20] and the proposed method, respectively. The corresponding ground truth images are shown in Figure 4(d) and Figure 5(d). A local magnification (×10) of the red square marked regions in the Figures 4(a–d) and 5(a–d) are also shown to highlight the corresponding images for their detailed visualization and interpretation. The salient features present in the magnified portions of the Figures 4(c) and 5(c) are enhanced significantly compared to the highlighted portions of Figures 4(a–b) and 5(a,b). Results show the close resemblance of the Figures 4(c) and 5(c) with their ground truth images shown in Figures 4(d) and 5(d), respectively. The values of PSNR, SSIM and FSIM of the reconstructed images using the proposed method improve by ≈ 1.23 dB, ≈ 0.03 and ≈ 0.01, respectively over the existing conventional CS [9, 19] and DL [20, 26] based methods as shown in Figures 4 and 5. Results support that the proposed approach enables better reconstruction compared to the other approaches.

Figures 6 and 7 depict the representative reconstruction of a chest X-ray test image to demonstrate further visual comparisons and more generalized behavior of the trained network. Figure 6(a–c) are the reconstructed images using the conventional CS based on GSR [9], BCS [19] and the proposed methods, respectively. Figure 7(a–c) represent the reconstructed images using the DL based DeepCNN [26], ConvNet [20] and the proposed method, respectively. The red marked regions of Figures 6(a-d) and 7(a–d) are magnified (10X) and presented above the corresponding images. As can be seen, the highlighted portion of the reconstructed images presented in the Figures 6(c) and 7(c) are almost indistinguishable from the highlighted portion of the ground truth images in the Figures 6(d) and 7(d), respectively. An improvement at

### Table 4 Hyper parameters used in the training

| Parameter name       | Value       |
|----------------------|-------------|
| Learning Rate, \( \alpha \) | 0.0012     |
| Momentum             | 0.0         |
| Dropout Rate         | 0.4         |
| Batch Size           | 16          |
| Optimizer used       | Gradient descent |

![Figure 3](image-url)  
Figure 3 Training progress evaluation of reconstruction network: reconstruction loss versus epoch
least by an amount of $\text{PSNR} \approx 0.64 \, \text{dB}, \text{SSIM index} \approx 0.03$ and $\text{FSIM index} \approx 0.01$ for the reconstructed images presented in Figures 4(c) and 7(c) compared to the Figures 4(a,b) and 7(a,b) ensure that the reconstruction quality is improved significantly. It is worth mentioning that all the images presented in this subsection are reconstructed from 10% measurements that obtained by the sensing network as compressed features.

This work also presents a quantitative comparison in terms of the PSNR, SSIM and FSIM values of the reconstructed images evaluated on different test dataset of brain [34], retina [32, 33] and chest X-ray [35] as reported in Table 5. The values reported are obtained by averaging 10 different run for the test dataset when the sampling ratio are used as 0.01, 0.05 and 0.10, respectively. To perform this study the reconstruction network defines the number of neurons in the 1st FC-layer as 512, 2460 and 4916, respectively to achieve the desired reduction factors, while reconstructing an image of size $128 \times 128 \times 3$. An improved reconstruction capability of the proposed method over the existing conventional CS based GSR [9], BCS [19] methods as well as the DL based DeepCNN [26], ConvNet [20] methods validate the effectiveness of the proposed method.

### 3.4.1 Reconstruction time comparison

This subsection presents a comparative study on the reconstruction time needed by the proposed and the existing GSR [9], BCS [19], DeepCNN [26] and ConvNet [20] methods as summarized

---

**Figure 4** Visual comparison: (a–c) are the reconstructed images from the 10% measurements for the methods GSR [9], BCS [19] and the proposed approach, respectively along with their local magnification; (d) the ground truth retina image and it’s local magnification

---

**Figure 5** Visual comparison: (a–c) are the reconstructed images from the 10% measurements for the methods DeepCNN [26], ConvNet [20] and the proposed approach, respectively along with their local magnification of the square marked regions; (d) the ground truth retina image and it’s local magnification
Figure 6 Visual comparison: (a–c) are the reconstructed images from the 10% measurements for the methods GSR [9], BCS [19] and the proposed approach, respectively along with their local magnification of the square marked regions; (d) the ground truth chest X-ray and it’s local magnification.

Figure 7 Visual comparison: (a–c) are the reconstructed images from the 10% measurements for the methods DeepCNN [26], ConvNet [20] and the proposed approach, respectively along with their local magnification of the square marked regions; (d) the ground truth chest X-ray image and it’s local magnification.

Table 5 Performance comparison: Average PSNR, SSIM and FSIM comparisons of the reconstructed images on brain [34], retina [32, 33] and chest X-ray [35] datasets.

| Data     | GSR [9] | BCS [19] | DeepCNN [26] | ConvNet [20] | PM       |
|----------|---------|----------|--------------|--------------|----------|
|          | PSNR    | SSIM     | FSIM         | PSNR         | SSIM     | FSIM     | PSNR    | SSIM     | FSIM     | PSNR    | SSIM     | FSIM     |
| Brain    | 0.01    | 25.57    | 0.77         | 0.87         | 26.04    | 0.77     | 0.86     | 28.44    | 0.80     | 0.88     | 28.53    | 0.82     | 0.89    | 29.40    | 0.85     | 0.93    |
| Retina   | 0.05    | 26.71    | 0.79         | 0.89         | 26.87    | 0.80     | 0.88     | 28.50    | 0.82     | 0.91     | 28.67    | 0.83     | 0.90    | 30.31    | 0.87     | 0.92    |
| Chest    | 0.05    | 28.32    | 0.80         | 0.89         | 28.85    | 0.81     | 0.89     | 31.27    | 0.82     | 0.91     | 31.20    | 0.82     | 0.92    | 33.08    | 0.88     | 0.93    |
| Brain    | 0.10    | 28.42    | 0.81         | 0.89         | 29.00    | 0.82     | 0.89     | 31.50    | 0.84     | 0.91     | 31.91    | 0.86     | 0.93    | 33.62    | 0.89     | 0.94    |
| Retina   | 0.10    | 30.05    | 0.83         | 0.90         | 31.44    | 0.84     | 0.90     | 33.22    | 0.87     | 0.95     | 34.16    | 0.88     | 0.94    | 35.51    | 0.90     | 0.95    |
| Chest    | 0.10    | 28.80    | 0.82         | 0.88         | 28.01    | 0.82     | 0.89     | 30.80    | 0.87     | 0.94     | 31.06    | 0.87     | 0.93    | 32.97    | 0.89     | 0.94    |
in Table 6. In the conventional CS method, the reconstruction time depends on the technique and the number of iterations used for the algorithm. This study considers 100 number of iterations for all the CS methods. The block size required to run the BCS [19] method is considered as $(32 \times 32)$. In the DL based methods, the reconstruction time depends on the total number of learnable parameters, platform and tools used. Therefore, the learnable parameters of all the DL methods along with the reconstruction time are reported as shown in Table 6. This comparative study considers the size of image as $(128 \times 128)$ for all the methods. As the parameters of FC layers depend on the size of the image to process, the reported total number of learnable parameters exclude the number of parameters used in the FC layer to maintain a fair comparison. This study shows that the reconstruction time required by the proposed method is a bit higher than the DL based DeepCNN [26] and ConvNet [20] methods due to having the larger number of parameters, whereas it is at least 86 times faster than the conventional CS methods [19].

## 4 CONCLUSIONS AND FUTURE WORKS

This paper proposes a DL based CS framework for a high dimensional image reconstruction from the highly undersampled measurements. A sensing network is used for capturing a high dimensional image through a hierarchical convolutional layers that reduce the dimension of the image progressively. The DL framework uses another convolutional network to reconstruct back the original image from the compressed features vector. The reconstruction network is trained in presence of a discriminator network that helps to enhance the learning capability of the reconstruction network by providing a feedback. The proposed framework enables the reconstruction of a signal through a simple linear operation instead of using time consuming non-linear iterative approach like the conventional CS. Simulation results show that the image reconstruction quality is improved significantly when measured in terms of PSNR, SSIM and FSIM at least by $1.5 \pm 0.63$ dB, $0.04 \pm 0.02$ and $0.02 \pm 0.02$, respectively over the existing methods [9, 19, 20, 26].

The proposed system model works fine for the types of images used in the training. However, the trained network doesn’t work well for the other types of images. The future work may be considered to form a generalized network for CS reconstruction irrespective of the image dataset used in the training.

---

### Table 6 Performance comparison: approximate number of parameters used and reconstruction time

| Method Used       | GSR [9] | BCS [19] | DeepCNN [26] | ConvNet [20] | Proposed Method |
|-------------------|---------|----------|--------------|--------------|-----------------|
| Total Parameters  | –       | –        | 1,86,700     | 57,400       | 2,69,500        |
| Recons. Time (Sec)| 242.47  | 117.30   | 1.21         | 0.26         | 1.34            |

---

### ORCID

Santi P. Maity 📲 https://orcid.org/0000-0002-1075-3829

### REFERENCES

1. Donoho, D.L.: Compressed Sensing. IEEE Trans. Inf. Theory 52(4), 1289–1306 (2006)
2. Baraniuk, R.G.: Compressive Sensing. IEEE Signal Process. Mag. 24(4), 118–120 (2007)
3. Poole, A.G., et al.: Three dimensional OCT images from retina and skin. Opt. Express 7(9), 292–298 (2000)
4. Herman, M.A., Strohmer, T.: High-resolution radar via compressed sensing. IEEE Trans. Signal Process. 57(6), 2275–2284 (2009)
5. Hardie, R.C., et al.: High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. Opt. Eng. 37(1), 247–261 (1998)
6. LeCun, Y., et al.: Deep learning. Nature 521(7553), 436 (2015)
7. Goodfellow, I., et al.: Deep Learning. MIT Press, Cambridge, MA 2016
8. Candès, E.J., Wakin, M.B.: An introduction to compressed sensing. IEEE Signal Process. Mag. 25(2), 21–30 (2008)
9. Zhang, J., et al.: Group-based sparse representation for image restoration. IEEE Trans. Image Process. 23(8), 3336–3351 (2014)
10. Li, X., et al.: Patch matching-based multitemporal group sparse representation for the missing information reconstruction of remote-sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9(8), 3629–3641 (2016)
11. Shi, W., et al.: Group-based sparse representation for low lighting image enhancement. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 4082–4086, IEEE, Piscataway, NJ (2016)
12. Li, X., et al.: Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning. IEEE Trans. Geosci. Remote Sens. 52(11), 7086–7098 (2014)
13. Shen, H., et al.: Compressed sensing-based inpainting of aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse Bayesian dictionary learning. IEEE Trans. Geosci. Remote Sens. 52(2), 894–906 (2015)
14. Esfahani, N., Aghagolzadeh, A.: Compressive sensing image restoration using adaptive curvelet thresholding and nonlocal sparse regularization. IEEE Trans. Image Process. 25(7), 3126–3140 (2016)
15. Tian, Z., et al.: Low-dose CT reconstruction via edge-preserving total variation regularization. Phys. Med. Biol. 56(18), 5949 (2011)
16. Candes, E.J.: Enhancing sparsity by reweighted l1 minimization. Journ. of Fourier Analysis and Applications 14(5–6), 877–905 (2008)
17. Daubechies, I.: Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math. 63(1), 1–38 (2010)
18. Mun, S., Fowler, J.E.: Block compressed sensing of images using directional transforms. In: Proceedings of the IEEE international conference on image processing (ICIP), pp. 3021–3024, IEEE, Piscataway, NJ (2009)
19. Unde, A.S., Deepthi, P.: Block compressive sensing: Individual and joint reconstruction of correlated images. J. Visual Commun. Image Represent. 44, 187–197 (2017)
20. Dong, C., et al.: Learning a deep convolutional network for image super-resolution. In: Proceedings of the European Conference on Computer Vision, pp. 184–199, Springer, Berlin (2014)
21. Quan, T.M., et al.: Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. IEEE Trans. Med. Imaging 37(6), 1488–1497 (2018)
22. Goodfellow, I., et al.: Generative adversarial nets. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 2672–2680. MIT Press, Cambridge, MA (2014)
23. Sun, L., et al.: A deep information sharing network for multi-contrast compressed sensing MRI reconstruction. IEEE Trans. Image Process. 28(12), 6141–6153 (2019)
24. Yaman, B., et al.: Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data. In: Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI), pp. 921–925, IEEE, Piscataway, NJ (2020)
25. Lee, D., et al.: Deep residual learning for compressed sensing MRI. In: Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI), pp. 15–18, IEEE, Piscataway, NJ (2017)
26. Guo, D., et al.: Speedy and accurate image super-resolution via deeply recursive CNN with skip connection and network in network. IET Image Proc. 13(7), 1201–1209 (2019)
27. Shi, W., et al.: Image compressed sensing using convolutional neural network. IEEE Trans. Image Process. 29, 375–388 (2019)
28. Kulkarni, K., et al.: Non-iterative reconstruction of images from compressively sensed measurements. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 449–458. IEEE, Piscataway, NJ (2016)
29. Lu, X., et al.: Convcsnet: A convolutional compressive sensing framework based on deep learning. arXiv:180110342 (2018)
30. Qiao, J., et al.: Image super-resolution using conditional generative adversarial network. IET Image Proc. 13(14), 2673–2679 (2019)
31. Mousavi, A., et al.: A deep learning approach to structured signal recovery. In: Proceedings of the Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1336–1343, IEEE, Piscataway, NJ (2015)
32. Staal, J., et al.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging 23(4), 501–509 (2004)
33. Hoover, A., et al.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging 19(3), 203–210 (2000)
34. IXI Brain MR dataset. Imperial College London. http://brain-development.org/ixi-dataset/
35. Mooney, P.: Chest X-ray image dataset. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia/
36. Hinton, G.E., et al.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv:12070580 (2012)
37. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
38. Zhang, L., et al.: FSIM: A feature similarity index for image quality assessment. IEEE Trans. Image Process. 20(8), 2378–2386 (2011)

How to cite this article: Islam SR, Maity SP, Ray AK. On learning based compressed sensing for high resolution image reconstruction. IET Image Process. 2021;15:393–404. https://doi.org/10.1049/ipr2.12029