Structure Preserving Compressive Sensing MRI Reconstruction using Generative Adversarial Networks

Puneesh Deora[1*]   Bhavya Vasudeva[1*]   Saumik Bhattacharya[2]   Pyari Mohan Pradhan[1]

[1]Dept. of ECE, IIT Roorkee, Uttarakhand, India
[2]Dept. of E&ECE, IIT Kharagpur, West Bengal, India

{pdeora,bvasudeva}@ec.iitr.ac.in   saumik@ece.iitkgp.ac.in   pmpradhan.fec@iitr.ac.in

Abstract

Compressive sensing magnetic resonance imaging (CS-MRI) accelerates the acquisition of MR images by breaking the Nyquist sampling limit. In this work, a novel generative adversarial network (GAN) based framework for CS-MRI reconstruction is proposed. Leveraging a combination of patch-based discriminator and structural similarity index based loss, our model focuses on preserving high frequency content as well as fine textural details in the reconstructed image. Dense and residual connections have been incorporated in a U-net based generator architecture to allow easier transfer of information as well as variable network length. We show that our algorithm outperforms state-of-the-art methods in terms of quality of reconstruction and robustness to noise. Also, the reconstruction time, which is of the order of milliseconds, makes it highly suitable for real-time clinical use.

1. Introduction

Magnetic resonance imaging (MRI) is a commonly used non-invasive medical imaging modality that provides soft tissue contrast of excellent quality as well as high resolution structural information. The most significant drawback of MRI is its long acquisition time as the raw data is acquired sequentially in the k-space which contains the spatial-frequency information. This slow imaging speed can cause patient discomfort, as well as introduce artefacts due to patient movement.

Compressive sensing (CS) [10] can be used to accelerate the MRI acquisition process by undersampling the k-space data. Reconstruction of CS-MRI is an ill-posed inverse problem [13]. Conventional CS-MRI frameworks assume prior information on the structure of MRI by making use of predefined sparsifying transforms such as the discrete wavelet transform, discrete cosine transform, etc. to
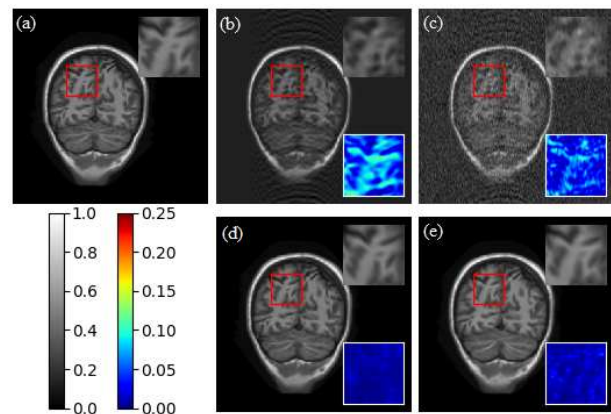
---
*Equal contribution

Figure 1. Our method takes zero-filled reconstruction (ZFR) of the undersampled image as input and generates the corresponding reconstructed image. This can essentially be viewed as de-aliasing the ZFR. Example reconstruction results when 30% data is retained. (a) Ground truth (GT), (b) ZFR of noise free image, (c) ZFR of image with 10% noise, (d) results of the proposed method for noise-free image, and (e) results of the proposed method for image with 10% noise. The top right inset indicates the zoomed in region of interest (ROI) corresponding to the red box, and the bottom right inset indicates the absolute difference between the ROI and the corresponding GT. The images are normalized between 0 and 1.

obtain the solution [22]. Instead of using predefined transforms, the sparse representation can be learnt from the data itself, i.e. dictionary learning (DLMRI) [26]. In [11], a different approach of alternating between solving the optimization problem for reconstruction and denoising the image using block matching 3D (BM3D) model is adopted. These frameworks however, suffer from the long computation time taken by iterative optimization processes [25] as well as the assumption of sparse signals [22], which might not be able to fully capture the fine details [21].

Bora *et al.* [5] have shown that instead of using the sparsity model, the CS signal can be recovered using pretrained generative models, where they use an iterative optimization

to obtain the reconstructed signal. Another deep learning based approach was introduced by Yang *et al.* [34], where a data flow graph is designed for alternating direction method of multipliers [6] to train the network (DeepADMM) for CS-MRI reconstruction. The inference phase takes a time similar to ADMM although the optimized parameters used are learned during the training process. A network architecture resembling a cascade of convolutional neural networks (CNNs) is proposed in [29] (DeepCascade) which aims to reconstruct dynamic sequences as well as independent frames of 2D MR images undersampled using Cartesian masks. The cascading network laid out resembles dictionary learning reconstruction approaches, where the proposed approach can be viewed as an extended version of DLMRI. In [19], the authors unroll a residual learning approach where they use a deep CNN to learn the aliasing artifacts in the undersampled image, and subtract the aliasing artifacts thus estimated to obtain the de-aliased output.

Recent works [33, 23] demonstrate the application of generative adversarial networks (GANs) [12] to reconstruct CS-MRI. In these works, the use of a large set of CS-MR images and their fully sampled counterparts for training the GAN model can facilitate the extraction of prior information required to solve the reconstruction problem [32]. The trained model is then used to obtain the reconstructed output for a new CS-MR image in a very short time. In [33], the authors propose a refinement learning based approach to obtain the de-aliased reconstructed MR image using a conditional GAN framework (DAGAN). Mardani *et al.* [23] (GANCS) use pixel-wise ℓ_1/ℓ_2 loss to train the generator and a least-squares GAN framework.

Many of the aforementioned works, including DeepCascade and DAGAN use ℓ_2 loss function in the pixel domain for training, which is known to give blurry and excessively smooth outputs. Minimizing the ℓ_2 or ℓ_1/ℓ_2 norm of the pixel-wise difference does result in a higher peak-signal-to-noise ratio (PSNR) of the reconstructed image, but it does not ensure good reconstruction of the structural details [31]. In terms of frequency, the use of pixel-wise difference based loss mainly focuses on preserving low frequency components and does not enforce good reconstruction of high frequency details. Moreover, these state-of-the-art methods use discriminators which consider the input in a global sense while classifying. This may not allow the discriminator to consider the fine high frequency textural details, which are of vital importance in the MR images. Although DAGAN uses a frequency domain ℓ_2 loss, it has the drawback of penalizing larger differences more, and allowing several smaller differences. This can yield a reconstructed output that looks similar to the ground truth but fails to preserve the finer details in the form of high frequency and structural content.

Contributions: To overcome these drawbacks, we incorporate the ℓ_1 norm of the pixel-wise difference in the generator loss function to avoid blurry reconstruction. In order to preserve the structural and textural details in the reconstructed image, we propose the use of a structural similarity (SSIM) index based loss to train the generator. Moreover, to ensure better reconstruction of high frequency content in the MR images, we propose the use of a patch-based discriminator. Further, we propose a novel generator architecture by incorporating residual in residual dense blocks (RRDBs) in a U-net based architecture to utilize the benefits of residual and dense connections. It is also known that the binary cross-entropy based adversarial loss, which has been used in most of the previous works, makes the training of GANs unstable. Therefore, in order to stabilize the training process, we incorporate the Wasserstein loss. The reconstructed images should be less sensitive to the noise level in the measurements, since hardware devices are always susceptible to noise. In order to make the reconstruction robust to noise, we propose the use of noisy images for data augmentation to train our GAN model. Fig. 1 shows an example of reconstruction results obtained by the proposed approach (described in section 2 and 3) on noise-free as well as images contaminated with noise.

2. Methodology

The acquisition model for the CS-MRI reconstruction problem in discrete domain can be described as:

$$u = \mathbf{G}y + \eta, \tag{1}$$

where $y \in C^{N^2}$ is a vector formed by the pixel values in the $N \times N$ desired image, $u \in C^M$ denotes the observation vector, and $\eta \in C^M$ is the noise vector. C denotes the set of complex numbers. The matrix $\mathbf{G}$ describes the process of random undersampling in the k-space. It is the product of an $N^2 \times N^2$ matrix $\mathbf{F}$, which computes the Fourier transform, and an $M \times N^2$ undersampling matrix $\mathbf{U}$. Given an observation vector u, the reconstruction problem is to find out the corresponding y, considering η to be a non-zero vector. We choose to find the solution to this reconstruction problem using a GAN model.

A GAN model comprises of a generator G and a discriminator D, where the generator tries to fool the discriminator by transforming input vector z to the distribution of true data y_{true}. On the other hand, the discriminator attempts to distinguish samples of y_{true} from generated samples $G(z)$. We incorporate the conditional GAN (cGAN) based framework [24] in our study. The model is conditioned on the aliased zero-filled reconstruction (ZFR) $x \in C^{N^2}$, given by $x = \mathbf{G}^H u$, where H denotes the Hermitian operator. Instead of using a binary cross-entropy based adversarial loss for training the cGAN model, we use the Wasserstein loss [2]. This helps in stabilizing the training process of stan-
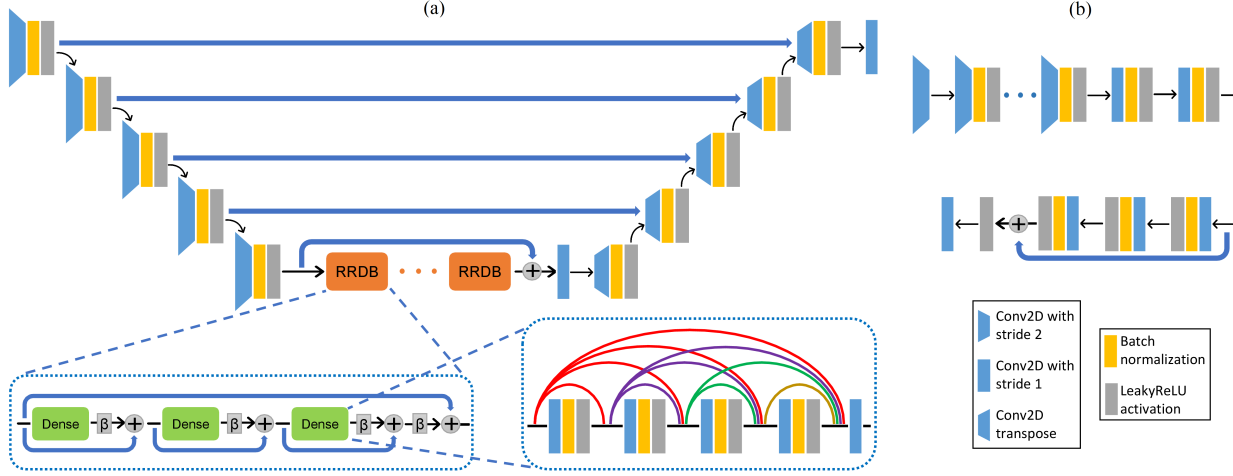
Figure 2. (a) Generator architecture and (b) discriminator architecture.

dard GANs, which suffer from saturation resulting in vanishing gradients. Mathematically, the cGAN model with the Wasserstein loss solves the following optimization problem:

$$\min_{G} \max_{D} V_{WGAN} = \mathbb{E}_{\mathbf{y} \sim p_y(\mathbf{y})}(D(y)) \\ - \mathbb{E}_{\mathbf{x} \sim p_x(\mathbf{x})}(D(G(x))), \quad (2)$$

where V_{WGAN} denotes the value function and $\mathbb{E}$ denotes the expectation over a batch of images. $p_y(\mathbf{y})$ and $p_x(\mathbf{x})$ denote the distribution of GT and ZFR images, respectively. The optimization problem is solved by alternating between p steps where discriminator (D) is optimized and a single step of generator (G) optimization. The loss function which is minimized while training the discriminator is given by:

$$L_{DIS} = \mathbb{E}_{\mathbf{x} \sim p_x(\mathbf{x})}(D(G(x))) - \mathbb{E}_{\mathbf{y} \sim p_y(\mathbf{y})}(D(y)). \quad (3)$$

The Lipschitz constraint is enforced by applying weight clipping on the weights of the discriminator [2].

Fig. 2 (a) shows the generator architecture of the proposed model. The architecture is based on a U-net [27], which consists of several encoders and corresponding decoders. Each encoder is in the form of a convolutional layer, which decreases the size and increases the number of feature maps. Each decoder consists of a transposed convolutional layer, to increase the size of the feature maps. In order to transfer the features of a particular size from the encoder to the corresponding decoder, skip connections are present. Instead of obtaining feature maps of size lower than $\frac{N}{32} \times \frac{N}{32}$ using more encoders (and decoders), the proposed architecture consists of RRDBs at the bottom of the U-net. The addition of RRDBs at the bottleneck layer helps in increasing the depth of the network which can enable learning of more complicated functions. Each RRDB [30] consists of dense blocks, as well as residual connections at two levels: across each dense block, and across all the dense blocks

in one RRDB, as shown in Fig. 2 (a). The output of each dense block is scaled by β before it is added to the identity mapping. Residual connections make the length of the network variable thereby making identity mappings easier to learn and avoid vanishing gradients in the shallower layers. Dense connections allow the transfer of feature maps to deeper layers, thus increasing the variety of accessible information. Just like residual connections, they also help in alleviating vanishing gradients. Moreover, their use reduces the number of parameters as compared to conventional convolutional networks, since the necessity to learn redundant information has been removed. Throughout this network, batch normalization (BN) and leaky rectified linear unit (ReLU) activation are applied after each convolutional layer. At the output, a hyperbolic tangent activation is used.

The discriminator is a CNN with 11 layers, as illustrated in Fig. 2 (b). Each layer consists of a convolutional layer, followed by BN and leaky ReLU activation. A patch-based discriminator [14] is incorporated in order to improve the preservation of high frequency details in the reconstructed output, since ℓ_1 norm of the pixel-wise difference (used as a loss function in this work) mainly focuses on preservation of low frequency components and does not enforce good reconstruction of high frequency details. This discriminator focuses on the local patches, tries to score each patch (size $m \times m$) of the image separately in an attempt to classify whether the patch is real or fake, and gives the average score as the final output.

In order to reduce the pixel-wise difference between the generated image and the corresponding ground truth (GT) image, a mean absolute error (MAE) based loss is incorporated while training the generator. It is given by:

$$L_{MAE} = \mathbb{E}(\|G(x) - y\|_1), \quad (4)$$

where $\| \cdot \|_1$ denotes the ℓ_1 norm. Since the human vision system is sensitive to structural distortions in images, it is important to preserve the structural information in MR images, which is crucial for clinical analysis. Moreover, ℓ_1 norm minimization of the pixel-wise difference does not enforce textural and structural correctness, which may lead to a reconstructed output of poor diagnostic quality. Super resolution is another well-known inverse problem that tries to interpolate both low frequency and high frequency components from a low resolution image. Inspired by previous works on super resolution [37], a mean SSIM (mSSIM) [35] based loss is incorporated in order to improve the reconstruction of fine textural details in the images. It is formulated as follows:

$$L_{mSSIM} = 1 - \mathbb{E}\left(\frac{1}{K} \sum_{j=1}^{K} SSIM(G_j(x), y_j) \right), \quad (5)$$

where K is the number of patches in the image, and SSIM is calculated as follows:

$$SSIM(u, v) = \frac{2\mu_u\mu_v + c_1}{\mu_u^2 + \mu_v^2 + c_1} \frac{2\sigma_{uv} + c_2}{\sigma_u^2 + \sigma_v^2 + c_2}, \quad (6)$$

where u and v represent two patches, and μ and σ denote the mean and variance, respectively. c_1 and c_2 are small constants to avoid division by zero.

The overall loss for training the generator is given by:

$$L_{GEN} = \alpha_1 L_{MAE} + \alpha_2 L_{mSSIM} - \alpha_3 \mathbb{E}(D(G(x))), \quad (7)$$

where α_1, α_2, and α_3 are the weighting factors for various loss terms.

3. Results and Discussion

3.1. Training settings

In this work, a 1-D Gaussian mask is used for undersampling the k-space. Since the ZFR x is complex valued, the real and imaginary components are concatenated and passed to the generator in the form of a two channel real valued input. The batch size is set as 32. The discriminator is updated three times before every generator update. The threshold for weight clipping is 0.05. The growth rate for the dense blocks is set as 32, β is 0.2, and 12 RRDBs are used. The number of filters in the last layer of each RRDB is 512. Adam optimizer [16] is used for training with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate is set as 10^{-4} for the generator and 2×10^{-4} for the discriminator. The weighting factors are $\alpha_1 = 20$, $\alpha_2 = 1$, and $\alpha_3 = 0.01$. The model [1] is implemented using Keras framework [8] with TensorFlow backend. For training, 4 NVIDIA GeForce GTX 1080 Ti GPUs are used, each having 11 GB RAM.

---

[1]The code is available at: `https://puneesh00.github.io/cs-mri/`

3.2. Data details

For the purpose of training and testing, two different datasets are used. We first evaluate our model on T-1 weighted MR images of brain from the MICCAI 2013 grand challenge dataset [18]. This is followed by another evaluation using MR images of knee (coronal view) from the MRNet dataset [4]. The images in both the datasets are of size 256×256. In order to make the reconstructed output robust to noise, data augmentation is carried out using images with 10% and 20% additive complex Gaussian noise in the k-space. To make the set of training images for the MICCAI dataset, 19 797 images are randomly taken from the training set of the aforementioned dataset. Out of these, noise is added to 6335 images, while the remaining 13 462 images are used without any noise. In addition, 990 images are chosen from the 13 462 noise-free images, and noise is added to them also, to get a total of 20 787 images for training. Among the noisy images, number of images with 10% and 20% noise is equal. Thus, the set contains 64.76% noise-free images, 30.48% noisy images whose corresponding noise-free images are not present in the training set, and 4.76% noisy images whose corresponding noise-free images are present in the training set. To make the set of training images for the MRNet dataset, a total of 12 500 images are taken from the training set, where the aforementioned ratio of noise-free, overlapping noisy images, and non-overlapping noisy images is maintained. For testing, 2000 images are chosen randomly from the test sets of the respective datasets. The tests are conducted in three stages: using noise-free images, using images with 10% noise added to them, and using images with 20% noise.

3.3. Results

Table 1 summarizes the quantitative results to study the effect of addition of various components to the model. These results are reported for images taken from the MICCAI dataset, in which 20% of the raw k-space samples are retained. For all the cases, the generator is trained with L_{GEN}. In the first case, the GAN model comprises of a U-net generator (without RRDBs) and a patch-based discriminator, with BN present throughout the network. It is trained with Wasserstein loss. In the subsequent cases, the use of RRDBs (without BN), followed by addition of BN to RRDBs results in significant improvement in PSNR of the reconstructed outputs corresponding to noise-free images. As mentioned in Section 3.1, the loss function used in training of all the networks takes the weight for L_{MAE} as 20 times the weight for L_{mSSIM}. Such a ratio might cause the model to focus more towards reducing the MAE. This results in a more consistent performance of PSNR as the training progresses. In the inference on noisy test images (both 10% and 20%), the PSNR and mSSIM have relatively less consistent performance as seen in the ablation studies.

Table 1. Ablation study of the model.

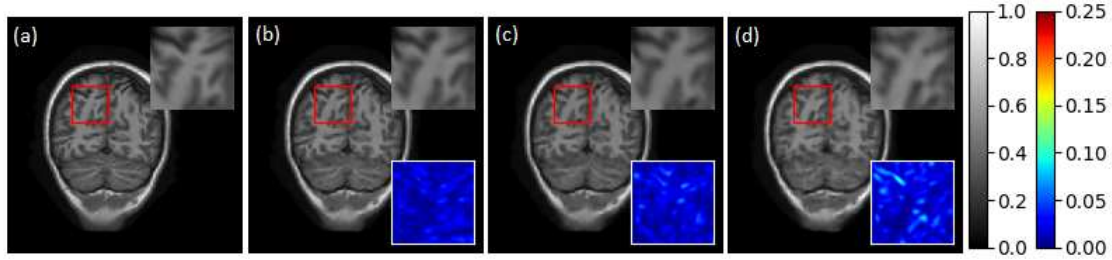| Network Settings	1st	2nd	3rd	4th	5th
U-net G + patch-based D	✓	✓	✓	✓	✓
RRDBs	✗	✓	✓	✓	✓
BN in RRDBs	✗	✗	✓	✓	✓
Data augmentation	✗	✗	✗	✓	✓
Wasserstein loss	✓	✓	✓	✗	✓
Images	PSNR (dB) / mSSIM				
Noise-free	40.45 / 0.9865	41.39 / 0.9810	41.88 / 0.9829	41.80 / 0.9820	42.31 / 0.9841
10% noise added	38.25 / 0.9641	38.03 / 0.9624	38.03 / 0.9620	39.55 / 0.9728	39.80 / 0.9751
20% noise added	33.98 / 0.9217	34.01 / 0.9210	33.78 / 0.9180	37.21 / 0.9576	37.56 / 0.9619

Figure 3. Reconstruction results of the proposed method for 20% undersampled images, taken from the MICCAI dataset. (a) GT, reconstruction results for (b) noise-free image, (c) image with 10% noise, and (d) image with 20% noise. The top right inset indicates the zoomed in ROI corresponding to the red box. The bottom right inset indicates the absolute difference between the ROI and the corresponding GT. The images are normalized between 0 and 1.

One possible reason for this observation might be the large number of nonlinearities present in the model, which give the ability to learn a highly complex function. As mentioned in [28], a highly complex function can have improved performance for the noise-free case at the cost of slightly increased sensitivity to noise as compared to its less complex counterparts. The use of data augmentation with noisy images, in the fifth case, results in significantly better quantitative results for the reconstruction of noisy images, as compared to the first three cases. This improves the robustness of the model. In the fourth case, we train the network with the conventional binary cross-entropy based adversarial loss instead of Wasserstein loss. On comparing this case with the fifth case, it is evident that the use of Wasserstein loss improves the training process. The settings of the fifth case are finalized and used for the subsequent results reported in this work.

The qualitative results of the proposed method are shown in Fig. 3 for 20% undersampled images taken from the MICCAI dataset. It can be seen that the proposed method is able to reconstruct the structural content in the image, including many fine details, successfully. This is also indicated by the quantitative results shown in Table 1. Also, the contrast of the reconstructed image looks very similar to that of the GT. The reconstruction results for noisy inputs, as well as their differences with the corresponding GT, indicate the robustness of the model.

Fig. 4 and Table 2 show the qualitative and quantitative comparison of the proposed method, respectively, with some state-of-the-art methods like DLMRI [26], DeepADMM [34], BM3D [11], and DAGAN [33]. These results are reported for images taken from the MICCAI dataset, in which 30% of the k-space data is retained. The comparison of the zoomed in ROI of the reconstructed outputs corresponding to the noise-free images, produced by the aforementioned methods, as well as the difference with the GT show that these methods are not able to fully preserve the structural content present in the GT. It can be seen that our method produces the least difference between the ROI and the corresponding GT. Even in the case of noisy images, our method is robust to the artifacts in the image as it produces a smooth background, similar to the GT, whereas other methods produce outputs with noisy artifacts as well as granularity. This can be seen in the results shown in the second and third row in Fig. 4. The artifacts are more visible in the background of the zoomed in ROI, whereas the granularity can be more easily seen in the greyish boundary surrounding the brain structure as well as in the black background. Moreover, the contrast is much better preserved in our reconstructed outputs as seen in the zoomed in ROI of all the three rows in Fig. 4.

The quantitative results also reinforce the effectiveness of the proposed method. Table 2 shows that both the PSNR and mSSIM for the proposed method are significantly better than the previous methods for noise-free as well as images with 10% and 20% noise. All the previous methods, with the exception of DAGAN, experience a significant decline in the PSNR and mSSIM values when their reconstruction
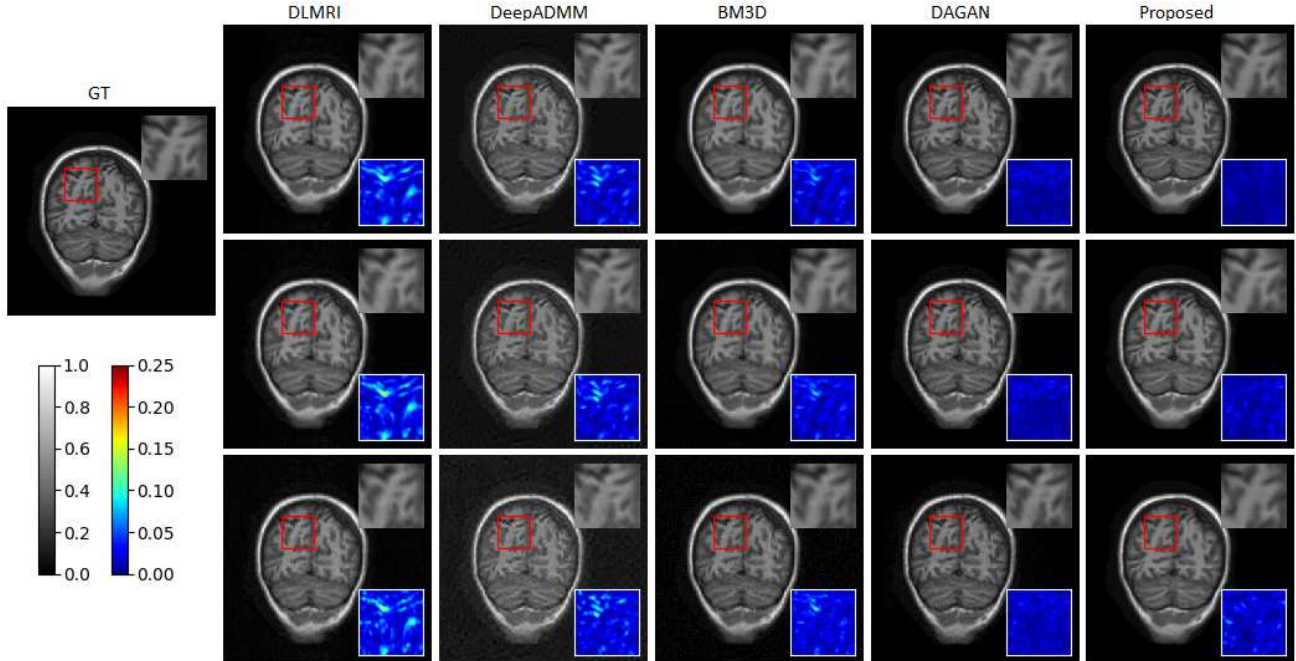
Figure 4. Qualitative results and comparison with previous methods for 30% undersampled images, taken from the MICCAI dataset. The first row shows reconstruction results for noise-free images, the second row shows reconstruction results for images with 10% noise, and the third row shows reconstruction results for images with 20% noise. The top right inset indicates the zoomed in ROI corresponding to the red box. The bottom right inset indicates the absolute difference between the ROI and the corresponding GT. The images are normalized between 0 and 1.

Table 2. Quantitative comparison with previous methods using MICCAI dataset.

Method	Noise-free images		10% noise added		20% noise added		Reconstruction/
	PSNR (dB)	mSSIM	PSNR (dB)	mSSIM	PSNR (dB)	mSSIM	Test time (s)
DLMRI[26]	37.405	0.8732	34.144	0.6140	31.564	0.4346	25.7732
DeepADMM[34]	41.545	0.8946	39.078	0.8105	35.373	0.6000	0.3135
BM3D[11]	42.521	0.9764	37.836	0.7317	33.657	0.4947	6.8230
DAGAN[33]	43.329	0.9860	42.006	0.9814	39.160	0.9619	**0.0063**
Proposed	**46.877**	**0.9943**	**42.338**	**0.9855**	**39.493**	**0.9740**	0.0091

results for noise-free and noisy images are compared. This proves that the reconstruction quality significantly deteriorates on addition of noise as these methods lack robustness. It is observed that the proposed method significantly outperforms the other methods in the noise-free setting, but the improvement in the noisy setting is less significant. As mentioned before, this might be the result of the large number of nonlinearities present in the model, which allow the learned function to be highly complex and obtain better performance for the noise-free case at the cost of slightly more sensitivity to noise [28]. However, the proposed augmentation technique increases the robustness of the model, as seen by the results presented in Table 2. Moreover, the reconstruction time of the proposed method is 9.06 ms per image, which can facilitate real-time reconstruction of MR images. DLMRI and BM3D have a much higher reconstruction time owing to the iterative fashion in which they obtain the output. On the other hand, GAN based approaches have

a reconstruction time of the order of milliseconds as the testing phase only involves a forward pass through the trained generator. For the generator model of DAGAN and the proposed approach, the FLOPs (total number of floating-point operations) are 197M and 314M, respectively. The corresponding FLOPS (floating-point operations per second) are 31.31G and 34.55G, which are calculated using the inference times listed in Table 2.

To demonstrate the generalization of the proposed approach, Table 3 and Fig. 5 show the quantitative and qualitative results for images taken from the MRNet dataset, in which 30% of the k-space data is retained, as well as the comparison with DAGAN [33], which obtained the closest results on the MICCAI dataset. From Table 3, it is observed that the proposed method outperforms DAGAN by a significant margin as it obtains better results for images with 20% noise than those obtained by DAGAN on noise-free images. It is also observed that the PSNR and mSSIM values ob-

tained for MRNet dataset are lower than those obtained for MICCAI dataset. One possible reason for this might be the larger black region present in the images in the MICCAI dataset, which lacks any details or structural information. Fig. 5 shows that the proposed method is able to obtain a reconstructed output of high quality, as the difference between the GT and the reconstructed image is very low.

Table 3. Quantitative results and comparison using MRNet dataset.

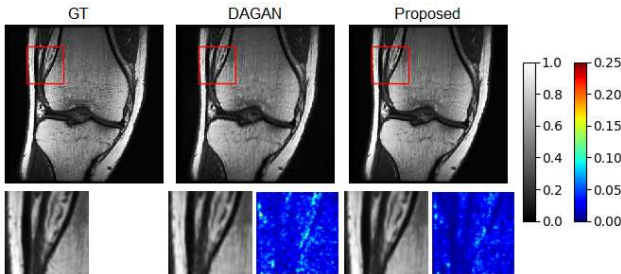| Method	PSNR (dB) / mSSIM		
	Noise-free images	10% noise added	20% noise added
DAGAN[33]	31.529 / 0.8754	30.452 / 0.8182	28.267 / 0.7098
Proposed	**34.823 / 0.9412**	**33.522 / 0.9167**	**32.034 / 0.8884**

Figure 5. Qualitative results and comparison for 30% undersampled image, taken from the MRNet dataset. These are the reconstruction results for noise-free images. The left inset indicates the zoomed in ROI corresponding to the red box. The right inset indicates the absolute difference between the ROI and the corresponding GT. The images are normalized between 0 and 1.
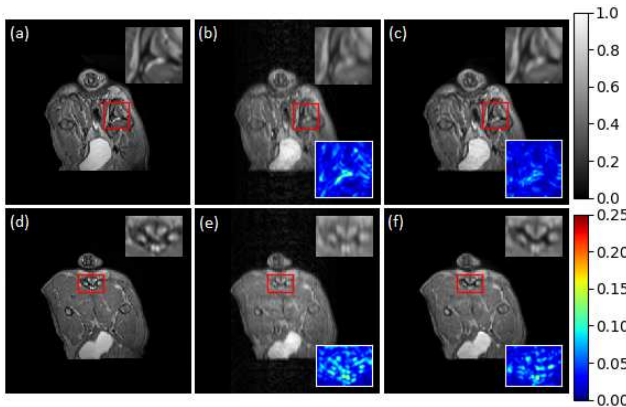

Figure 6. Results of zero-shot inference. (a,d) GT, (b,e) ZFR, (c,f) reconstruction results for noise-free image. The top right inset indicates the zoomed in ROI corresponding to the red box. The bottom right inset indicates the absolute difference between the ROI and the corresponding GT. The images are normalized between 0 and 1.

We also tested the model trained on MR images of brain from the MICCAI dataset to reconstruct MR images of canine legs from the MICCAI 2013 challenge. Fig. 6 shows the results of this zero-shot inference for images in which

20% of the k-space data is retained. Though no images of canine legs were used for training, the model is able to faithfully reconstruct most of the structural content, and is able to achieve average PSNR and mSSIM values of 41.28 dB and 0.9788, respectively, for 2000 test images.

Further, we performed the zero-shot inference of the model trained on 30% undersampled MR images of knee from the MRNet dataset to reconstruct MR images of canine legs from the MICCAI 2013 challenge. It is able to achieve average PSNR and mSSIM values of 43.79 dB and 0.9883, respectively, for 2000 test images.

Potential hallucination by GANs: Conventional GAN training techniques may suffer from hallucination of details which could potentially be harmful for image diagnosis. The proposed scheme tries to control the hallucination of details by the use of pixel-wise MAE loss as well as the mSSIM based loss in the image domain, both of which try to ensure that the generated image is close to the GT. L_{MAE} tries to make sure that the low frequency details are of the generated output as closely aligned to the ground truth, whereas L_{mSSIM} focuses on increasing the similarity of generated and GT in terms of structural details. The zero-shot inference is also helpful in pointing out that the proposed GAN model has shown no sign of hallucination on data samples taken from a distribution that is different from the training distribution.

Additional Experiment for Super Resolution: To understand the usability of the proposed model in other computer vision applications, we take super resolution (SR) as a vision task, for which the model is not optimized, and evaluate its performance on some commonly used datasets for this task. All the hyperparameter values mentioned in section 3.1 are maintained for this experiment, except α_1, which is set as 30. For training, patches of size 192×192 were used from the images present in the DIV2K [1] and the Flickr2K datasets. These are the high resolution (HR) or the GT images (y). This variable image size is supported by the fully convolutional architecture of the generator as well as the discriminator, i.e. neither of the networks involve the use of dense layers. The patch-based discriminator, which classifies sections of the image and not the entire image, also allows the image size to be variable. For the task of $4\times$ super resolution, the corresponding low resolution (LR) images of size 48×48 are used. These are obtained using bicubic downsampling, which is a widely used degradation model. As the size of the input and the output images is the same in our framework, we use the images obtained using bicubic interpolation on the low resolution images as the input for the model (x). For the purpose of testing, we use Set5 [3] and Set14 [36] datasets. The quantitative results of our model as well as the comparison with previous methods for $4\times$ super resolution are presented in Table 4. The PSNR and mSSIM values are calculated by considering only the Y

channel of the images, after converting them from RGB to YCbCr colorspace, as mentioned in several previous works on super resolution. The qualitative results obtained using our approach are illustrated in Fig. 7 and Fig. 8. Although the proposed framework is optimized for the task of CS-MRI reconstruction, it gives satisfactory performance for the super resolution task as well.

Table 4. Quantitative results and comparison with previous methods for the SR experiment.

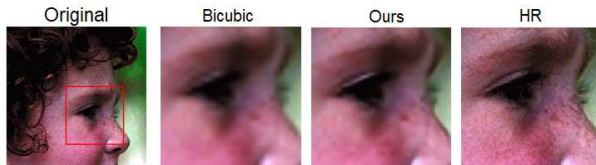| Method	PSNR (dB) / mSSIM	
	Set5	Set14
Bicubic	28.42 / 0.8104	26.00 / 0.7027
SRCNN[7]	30.48 / 0.8628	27.50 / 0.7513
VDSR[15]	31.35 / 0.8830	28.02 / 0.7680
FSRCNN[9]	30.72 / 0.8660	27.61 / 0.7550
LapSRN[17]	31.54 / 0.8850	28.19 / 0.7720
EDSR[20]	32.46 / 0.8968	28.80 / 0.7876
Ours	31.63 / 0.8982	27.62 / 0.7784

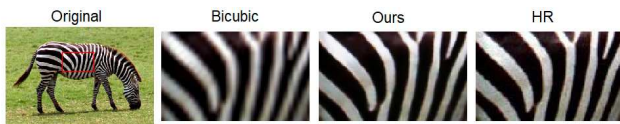Figure 7. Qualitative results of the SR experiment on an image from Set5.

Figure 8. Qualitative results of the SR experiment on an image from Set14.

4. Conclusion

In this paper, a novel GAN based framework has been utilized for CS-MRI reconstruction. The use of RRDBs in a U-net based generator architecture increases the amount of information available. In order to preserve the high frequency content as well as the structural details in the reconstructed output, a patch-based discriminator and structural similarity based loss have been incorporated. The use of noisy images during training makes the reconstruction results highly robust to noise. The proposed method is able to outperform the state-of-the-art methods, while maintaining the feasibility of real-time reconstruction. In future, we plan to analyze the performance of the proposed model for different k-space sampling patterns. In order to improve the reconstruction time, we plan to work on lightweight architectures. Further work may be carried out on devising regularization terms that help to preserve the finest of details in the reconstructed output.

References

[1] E. Agustsson and R. Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 214–223, International Convention Centre, Sydney, Australia, 06–11 Aug 2017.

[3] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012.

[4] N. Bien, P. Rajpurkar, R. L. Ball, J. Irvin, A. Park, E. Jones, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Medicine*, 15, 2018.

[5] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 537–546, 2017.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[7] K. He C. Dong, C. C. Loy and X. Tang. Learning a deep convolutional network for image super-resolution. In *The European Conference on Computer Vision (ECCV)*, page 184–199, 2014.

[8] F. Chollet et al. Keras. https://keras.io, 2015.

[9] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *The European Conference on Computer Vision (ECCV)*, 2016.

[10] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[11] E. M. Eksioglu. Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI. *Journal of Mathematical Imaging and Vision*, 56(3):430–440, Nov 2016.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.

[13] E. Herrholz and G. Teschke. Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems. *Inverse Problems*, 26(12):125012, nov 2010.

[14] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, July 2017.

[15] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.

[16] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, Conference Track Proceedings*, 2015.

[17] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[18] B. Landman and S. Warfield (Eds.). 2013 Diencephalon standard challenge.

[19] D. Lee, J. Yoo, and J. C. Ye. Deep residual learning for compressed sensing MRI. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 15–18, April 2017.

[20] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017.

[21] Y. Liu, J. F. Cai, Z. Zhan, D. Guo, J. Ye, Z. Chen, and X. Qu. Balanced sparse model for tight frames in compressed sensing magnetic resonance imaging. *PLOS ONE*, 10(4):1–19, 2015.

[22] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[23] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Transactions on Medical Imaging*, 38(1):167–179, Jan 2019.

[24] M. Mirza and S. Osindero. Conditional generative adversarial nets. *ArXiv*, abs/1411.1784, 2014.

[25] X. Qu, W. Zhang, D. Guo, C. Cai, S. Cai, and Z. Chen. Iterative thresholding compressed sensing MRI based on contourlet transform. *Inverse Problems in Science and Engineering*, 18(6):737–758, 2010.

[26] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041, May 2011.

[27] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597, 2015.

[28] P. Roy, S. Ghosh, S. Bhattacharya, and U. Pal. Effects of degradations on deep neural network architectures. *ArXiv*, abs/1807.10108, 2018.

[29] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, Feb 2018.

[30] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *The European Conference on Computer Vision (ECCV) Workshops*, pages 63–79, September 2018.

[31] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, Jan 2009.

[32] S. Xu, S. Zeng, and J. Romberg. Fast compressive sensing recovery using generative models with structured latent variables. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2967–2971, May 2019.

[33] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, June 2018.

[34] Y. Yang, J. Sun, H. Li, and Z. Xu. Deep ADMM-Net for compressive sensing MRI. In *Advances in Neural Information Processing Systems 29*, pages 10–18. Curran Associates, Inc., 2016.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[36] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *Proceedings of the 7th International Conference on Curves and Surfaces*, page 711–730, Berlin, Heidelberg, 2010. Springer-Verlag.

[37] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, March 2017.