2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

IMAGE QUALITY AFFECTS DEEP LEARNING RECONSTRUCTION OF MRI

Haris Jeelani[1], Jonathan Martin[1], Francis Vasquez[2], Michael Salerno[2], Daniel S. Weller[1]
[1]Electrical and Computer, [2] Biomedical, Engineering, University of Virginia, Charlottesville, USA

ABSTRACT

The magnetic resonance imaging (MRI) process is susceptible to a wide range of artifacts caused by various sources. In some cases, artifacts might be confused with pathology. In addition, state-of-the-art dynamic MR reconstruction algorithms are iterative in nature, causing longer reconstruction times. Recently, deep learning has been applied to MRI reconstruction and produces high quality images at high acceleration rates. Since deep learning highly depends on training data, the quality of training images must not be ignored. This article demonstrates how noisy images in the training data affect the quality of MR reconstruction. The proposed method modifies the loss function of the neural network to prefer higher quality target images by using a weighted loss function. In this paper mean squared error loss is used, but the approach can be extended to other types of loss function. Using still frames from cardiac MRI's, this approach is compared to existing approaches that discard noisy training data or ignore these quality differences. Even a basic weighting strategy improves the deep learning reconstruction quality over such methods.

Index Terms— magnetic resonance imaging, image reconstruction, Deep Learning, Image Quality assessment.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a powerful noninvasive imaging modality known for providing high spatial resolution diagnostic images without using ionizing radiation. MR images are acquired using various types of pulse sequences on different types of scanners. There have been significant developments in both static and dynamic MR image acquisition techniques over the past few years due to improvements in scanner hardware such as multicoil parallel imaging, increased field strength and beyond. Nevertheless, several artifacts can arise during accelerated image acquisition. For example, in parallel imaging, the distinctness of each coil's sensitivity behavior and suboptimal coil geometry can lead [1] to spatially varying noise in reconstructed images. In addition, dynamic MR images suffer from artifacts due to longer acquisition times and patient motion. Particularly in Cardiac MR (CMR) it is difficult to achieve high spatiotemporal resolution due to variation in breath holding times and heart motion. In perfusion MRI, the temporal dynamics have to be captured before the contrast agent washes out of the heart. These challenges have motivated the need for undersampled acquisition of CMR images. But, aliasing artifacts arise due to undersampling of k-space data followed by constrained reconstruction to obtain high quality reconstruction. Several state-of-the-art constrained reconstruction algorithms have been proposed based on sparsity [2] and low-rank [3] techniques. There are also data-adaptive techniques that learn the patch-based dictionary [4] directly using the sampled k-space data and adapt to the specific image instance. However, reconstruction times of nearly all constrained reconstruction algorithms are long because of their iterative nature.

Recently, deep neural network architectures have been applied [5] to inverse imaging and are shown to outperform total variation regularized reconstruction. They have also been applied in the form of stacked auto-encoder [6] and convolutional neural network (CNN) [7] to remove aliasing artifacts from dynamic MR images. In [7], a CNN performed better than a dictionary learning (DL) method [4] in terms of image quality and online processing time. CNNs and DL methods usually work in the image domain and are highly dependent on training data; therefore the quality of training images should be high. While [8] demonstrates how image quality affects performance of image classification task, limited research analyzes the effect of training CNNs with diagnostic images of varying quality on image reconstruction and also with images containing pathology. In this work, we use a similar network architecture as described in [7] to demonstrate how MR image reconstruction quality is affected by training the network with images of different image quality. Later we argue that the quality of the reconstruction can be improved by using a weighted loss function.

2. METHOD

The neural network architecture consists of three cascaded units of convolutional layer blocks and data consistency (DC) layers as shown in Figure 1. Each block of the convolutional layer is denoted as CNN_i, where 'i' signifies the position of the layer from the input, and is three layers deep with residual connections [9].The output of each convolutional block $\boldsymbol{x_{cnn}}$ is followed by a data consistency layer DC_i. Incorporating DC layers is important to prevent the network from changing the acquired k-space measurements $\widehat{\boldsymbol{x_u}}$ as the number of layers increases. To enforce consistency in k-space, the DC layer has a Fourier operator ($\boldsymbol{F}$) and a 2-norm-based data fit term ($\boldsymbol{f_{DC}}$) followed by an inverse Fourier operator ($\boldsymbol{F^{-1}}$). More details about the network can be found in [7]. The aliased input image $\boldsymbol{x_u}$ is trained against its unaliased version (the

target image). The network had two channels: one for real and another for imaginary values.

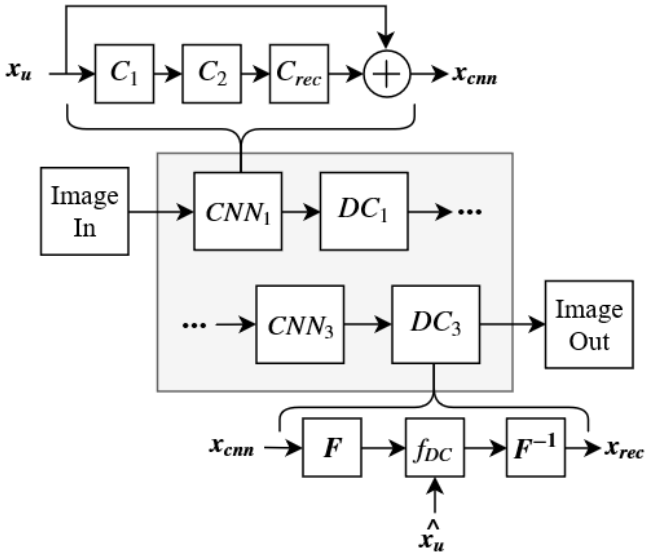

Figure 1: Convolutional neural network architecture with cascaded convolutional and data consistency layer.

The prediction of the network is denoted as $\boldsymbol{x_{rec}}$. The network was trained for 15 epochs with the Adam optimization method [10]. 'ImageIn' and 'ImageOut' represent input and output test images, respectively.

In order to make the network quality aware, we manually weight noisy images with a lower weighting and clean images with a higher weighting. The weights range between 0 (lowest) to 1 (highest). We use a weighted mean squared difference as our loss function $'L'$ to force the network to ignore the noisy predictions. We define the loss function for images of size (M=128, N=128) as:

$$L = \frac{\sum_{i=1}^{batchsize=20} w_i \sum_{j=1}^{M=128} \sum_{k=1}^{N=128} \left(|p_{ijk}| - t_{ijk}\right)^2}{\sum_{i=1}^{batchsize} w_i * M * N}$$

Here, $'w_i'$ is the weight of the i^{th} target image of size M x N. The $'p'_{ijk} \in \mathbb{C}$ and $'t'_{ijk} \in \mathbb{R}$ denote the pixel values of i^{th} prediction and target image respectively at positions j and k. As we use retrospectively acquired patient data (see Section 3), we possess only the magnitude images for these data, so we use the absolute value of the prediction to compute the squared difference.

3. EXPERIMENTS

Cine series is chosen as convenient sources of multiple images in order to form training and testing sets. The dataset consists of 14 short axis slices (each containing 25 cardiac phases) of cine videos acquired from two healthy subjects using a 3T scanner at the U.Va. Hospital. A standard protocol approved by the hospital IRB was followed and data were collected during sufficiently many heartbeats to reconstruct images without aliasing for all cardiac phases and slice locations. To conduct experiments, we create five training datasets from subject-1 by varying the number of noisy images. To degrade the image quality, zero mean Gaussian noise of variance up to 0.005 is individually added to the appropriate proportion of clean subject-1 data. Adding such noise approximates the expected degradation in quality from using accelerated sequences or using images from 1.5T scanners more common in CMR studies. The percentage of noisy images in training set is increased from 0% to 100% in the interval of 25%, giving rise to five datasets: N0, N25, N50, N75, and N100, respectively. For undersampling the data (retrospectively calculated from the magnitude images), we use a binary Cartesian mask with randomly varying phase encodes and with more lines near the center of k-space than the periphery. The aliased images are generated by taking 2D Inverse Fast Fourier Transform (IFFT) on zero-filled undersampled k-space. The testing set, from subject-2, remains clean and is subsampled in retrospectively generated k-space the same way. After generating datasets, we train the neural network for each dataset separately using the Lasagne [11] toolbox. The input to the network is the aliased image and the output is the un-aliased reconstructed image. The network is trained using images from subject-1 and is tested on subject-2 data. Note that the reconstruction is a still-frame reconstruction and not a dynamic one due to limited training data.

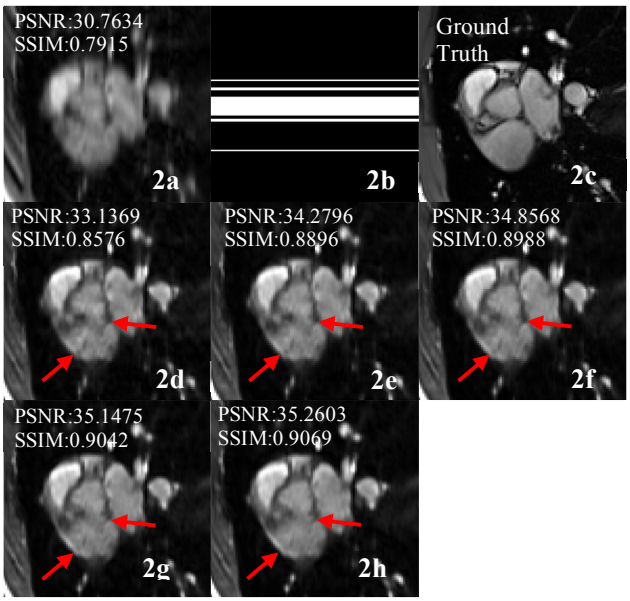

Figure 2a is the sample-aliased test input image-1. Figure 2b is the undersampling mask. Figure 2c is the ground truth. Figures 2d-h are the images reconstructed using N100-N0 training datasets with conventional (unweighted) training.

The following measures are taken to maintain consistency between experiments: the same random sampling mask (8x) is generated just once and is used to create noisy versions of the dataset, and the test set, batch size (20), learning rate (0.01), depth of cascade (up to CNN_3, DC_3) are the same for

all experiments. Although the individual layer weight initializer [12] is the same, the weights may be initialized differently for different experiments due to the functionality of the weight initialization.

In experiment-1, we demonstrate the effect of training the network with bad quality data on the reconstructed image. In experiment-2, we show how discarding or weighting the noisy images can affect the reconstructions. To validate the results of experiments we calculate PSNR and SSIM [13] around the heart and also perform visual assessment. Multiple repetitions of the experiments are performed to check the consistency of the results.

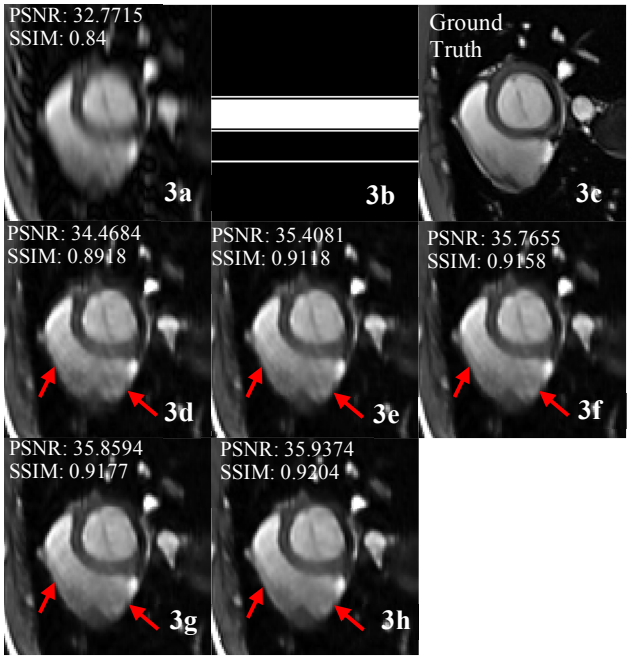

Figure 3a is the sample-aliased test input image-2. Figure 3b is the undersampling mask. Figure 3c is the ground truth. Figures 3d-h are the images reconstructed using N100-N0 training datasets with conventional (unweighted) training.

	N100	N75	N50	N25	N0
PSNR*	33.1603	33.8514	34.0635	34.2620	34.4194
SSIM*	0.8650	0.8872	0.8931	0.8968	0.9014

Table 1: Average PSNR* in dB and SSIM* values of five datasets created from subject-1. The '*' indicates metrics were computed on 80x80 cropped images to include only heart, as this region is of paramount importance for CMR.

4. DISCUSSION

Figure 2 shows the results of experiment-1 for two test images. The top left row in Figure 2a is the aliased input, and as we sequentially move along from Figure 2d-2h, we observe the quality of reconstructions increases as we decrease the number of noisy images in our training data. Figure 2c is the ground truth. Although reconstructions do not contain noise in them, artifacts can be seen clearly near the regions designated by arrows. Experiment-1 clearly demonstrates the effect of having the network trained on varying quality of images. We perform experiment-2 on two datasets: N25 and N50 assuming the amount of bad quality images in the training set would not reach beyond 50%. Upon training the network on both datasets, we find that on average the weighted loss function performs better than the non-weighted loss function, dropping or including noisy images. Although the numeric differences are small, the improvement is consistent: we do not find any case where including noisy images achieves a higher quality reconstruction than using the weighted loss on noisy images. In a majority of cases, we find removing noisy images causes blurred reconstructions, and is detrimental towards removing aliasing from images. The training time was around 3 minutes for each epoch on a Macbook Pro, and we observed the test loss was decreasing in the interval of $1e^{-5}$ after 10 epochs in every experiment, so we ran the experiments for 15 epochs. The average reconstruction time for an image in a batch size of 20 was around 250ms. This is significantly faster than compressed sensing reconstruction with a nonlinear conjugate gradient algorithm, which after choosing the right regularization parameters takes about a minute to reconstruct an image.

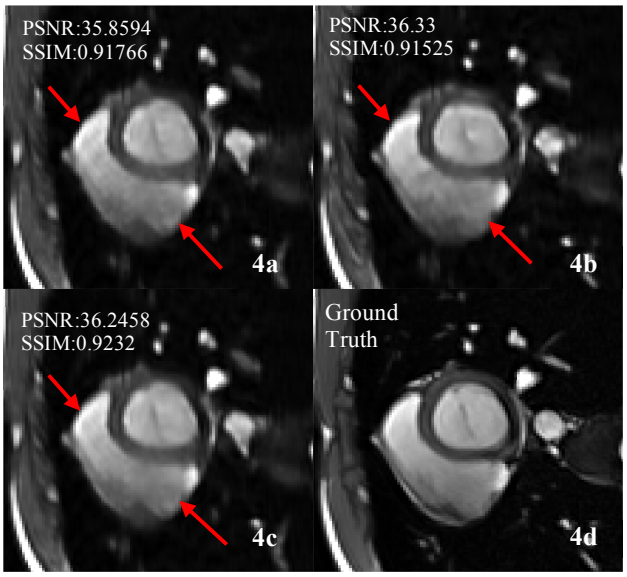

Figure 4 compares non-weighted and weighted loss reconstructions of a network trained using the N25 dataset. Figure 4a is the reconstruction of test input image-2 using a non-weighted loss and including noisy images. Figure 4b is reconstructed with noisy images removed. Figure 4c is the reconstruction using a weighted loss function.

Although differences are not significant, we expect to observe larger differences when we run these experiments over larger datasets, and use cross validation to choose the right set of hyperparameters. We also plan to include images with pathology and other artifacts as well. Selection of appropriate k-space sampling mask for neural network training is also a topic of further research. Since our approach

only modifies the loss function, it can be applied to any type deep learning network architecture.

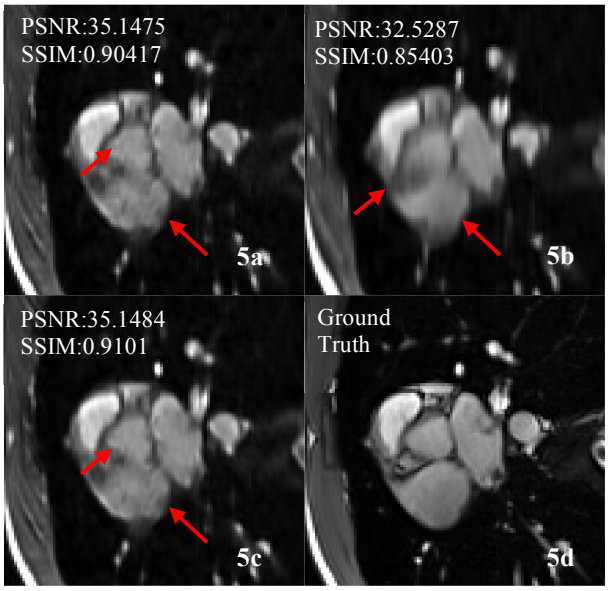

Figure 5 is similar to Figure 4, except that the test image is different (i.e., test image-1).

	Non-weighted loss & noisy images included	Non-weighted loss & noisy images dropped	Weighted loss & noisy included
PSNR*	34.4764	34.4107	34.6729
SSIM*	0.8951	0.8934	0.9009

Table 2: Average PSNR* in dB and SSIM* values for a network trained using 'N25' dataset. Two sample reconstructions are shown in Fig 4 and Fig 5 respectively.

	Non-weighted loss & noisy images included	Non-weighted loss & noisy images dropped	Weighted loss & noisy included
PSNR*	34.2666	34.1412	34.4655
SSIM*	0.8911	0.8875	0.8959

Table 3: Average PSNR* in dB and SSIM* values for a network trained on 'N50' dataset. Sample images are not shown for this table.

5. CONCLUSION

In this work, we evaluated the quality of MR images reconstructed by a deep neural network with datasets of varying image quality. Results from experiments show the quality of image reconstruction decreases if variable-quality training data are used. We also demonstrated that weighting the noisy training images with a lower weighting in the weighted loss function can improve the reconstructed test image quality with such training data.

Acknowledgements

The research team was supported by The Thomas F. and Kate Miller Jeffress Memorial Trust, Bank of America, Trustee.

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

6. REFERENCES

[1] P. Noël, R. Bammer, C. Reinhold, and M. A. Haider, "Parallel Imaging Artifacts in Body Magnetic Resonance Imaging," *Canadian Association of Radiologists Journal*, vol. 60, no. 2, pp. 91–98, Apr. 2009.

[2] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn Reson Med*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[3] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated Dynamic MRI Exploiting Sparsity and Low-Rank Structure: k-t SLR," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1042–1054, May 2011.

[4] S. Ravishankar and Y. Bresler, "MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, May 2011.

[5] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Trans Image Process*, Jun. 2017.

[6] A. Majumdar, "Real-time Dynamic MRI Reconstruction using Stacked Denoising Autoencoder," *CoRR*, vol. abs/1503.06383, 2015.

[7] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction," *IEEE Trans Med Imaging*, Oct. 2017, in press.

[8] S. F. Dodge and L. J. Karam, "Understanding How Image Quality Affects Deep Neural Networks," Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, 2016, pp. 1-6.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[10] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference for Learning Representations*, San Diego, 2015.

[11] S. Dieleman *et al.*, *Lasagne: First release.* 2015.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans Image Process*, vol. 13, no. 4, pp. 600–612, Apr. 2004.