

# UC San Diego

## UC San Diego Previously Published Works

### Title

Incorporating prior knowledge via volumetric deep residual network to optimize the reconstruction of sparsely sampled MRI

### Permalink

<https://escholarship.org/uc/item/6zz6690w>

### Authors

Wu, Yan  
Ma, Yajun  
Capaldi, Dante Pietro  
et al.

### Publication Date

2020-02-01

### DOI

10.1016/j.mri.2019.03.012

Peer reviewed



## Incorporating Prior Knowledge via Volumetric Deep Residual Network to Optimize the Reconstruction of Sparsely Sampled MRI

Yan Wu<sup>1</sup>, Yajun Ma<sup>2</sup>, Dante Pietro Capaldi<sup>1</sup>, Jing Liu<sup>3</sup>, Wei Zhao<sup>1</sup>, Jiang Du<sup>2</sup>, Lei Xing<sup>1,\*</sup>

<sup>1</sup>Radiation Oncology Department, Stanford University, Stanford, California, USA 94305

<sup>2</sup>Radiology Department, University of California San Diego, La Jolla, California, USA 92093

<sup>3</sup>Radiology Department, University of California San Francisco, San Francisco, California, USA 94107

### Abstract

For sparse sampling that accelerates magnetic resonance (MR) image acquisition, non-linear reconstruction algorithms have been developed, which incorporated patient specific *a priori* information. More generic *a priori* information could be acquired via deep learning and utilized for image reconstruction. In this study, we developed a volumetric hierarchical deep residual convolutional neural network, referred to as T-Net, to provide a data-driven end-to-end mapping from sparsely sampled MR images to fully sampled MR images, where cartilage MR images were acquired using an Ultra-short TE sequence and retrospectively undersampled using pseudo-random Cartesian and radial acquisition schemes. The network had a hierarchical architecture that promoted the sparsity of feature maps and increased the receptive field, which were valuable for signal synthesis and artifact suppression. Relatively dense local connections and global shortcuts were established to facilitate residual learning and compensate for details lost in hierarchical processing. Additionally, volumetric processing was adopted to fully exploit spatial continuity in three-dimensional space. Data consistency was further enforced. The network was trained with 336 three-dimensional images (each consisting of 32 slices) and tested by 24 images. The incorporation of *a priori* information acquired via deep learning facilitated high acceleration factors (as high as 8) while maintaining high image fidelity (quantitatively evaluated using the structural similarity index measurement). The proposed T-Net had an improved performance as compared to several state-of-the-art networks.

### Introduction

Sparse sampling and non-linear reconstruction algorithms have been extensively investigated to accelerate the acquisition of magnetic resonance (MR) images. In compressed sensing

\*corresponding author - Lei Xing, lei@stanford.edu 875 Blake Wilbur Drive Room G204, Stanford, California 94305.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(CS) [1], image sparsity has been explored in different domains (e.g. wavelet [1–3] and dictionary-based [4, 5] domains). It has also been applied with parallel imaging [6, 7]. Relatively high signal-to-noise ratio (SNR) as well as overall reduction in scan time have been achieved using CS [1], but unfortunately, difficulty in selecting parameters to optimize performance has limited the use of CS for real time application in routine clinical practice [8–10]. Meanwhile, patient specific *a priori* information has been incorporated into MR image reconstruction algorithms. Previous studies using *a priori* information have investigated the use of the support of signal [11, 12], the joint reconstruction of multi-contrast images [13] as well as the joint reconstruction of multiple time frame images in a dynamic sequence [13–15]. The success of these methods has strongly provided evidence of the value of integrating *a priori* information.

More recently, non-patient specific *a priori* information has been acquired via deep learning and utilized for image reconstruction. In fact, deep learning [16] has led to a flood of breakthroughs in image processing [17, 18], which is changing the landscape of medical physics [19, 20]. In some pilot studies on MRI reconstruction, deep learned *a priori* was incorporated into the framework of model based reconstruction [21–23], where neural networks were trained to predict fully sampled images that would be used as the initial image in compressed sensing [21], or to estimate the optimal value of parameters defined in compressed sensing [22, 23]. Alternatively, neural networks that performed end-to-end optimization were likely to provide better solutions, where data consistency enforcement and image sparsification were integrated [24–29]. In this way, the sparse representation was learned in the joint optimization instead of being obtained via a transform based on standard basis functions. In this study, we propose a volumetric hierarchical deep residual convolutional neural network framework, namely T-Net, to provide an end-to-end mapping for improved MRI reconstruction, where the flexibility of MR data sampling schemes was investigated via simulation, and the degree of acceleration were explored. We incorporated recent developments in deep learning techniques into the network design. The hierarchical network architecture enabled the extraction of feature maps at different scales, resulting in higher degree of sparsity as well as increased receptive field, which provided a wide range of context information for signal synthesis and artifact suppression. Relatively dense local shortcut connections were established to facilitate residual learning, whereas global shortcut connections were employed for compensating details lost in down-sampling. Additionally, volumetric processing was adopted to fully exploit spatial continuity in three-dimensional space. To further enhance data consistency, k-space data of the predicted images were replaced by the original measurements. While the proposed framework was generic for image reconstruction, in this study it was mainly applied on cartilage MRI acquired using an ultra-short echo-time (UTE) sequence and retrospectively undersampled in various trajectories.

## Methods

A volumetric hierarchical deep residual convolutional neural network (T-Net) was proposed to establish an optimal end-to-end mapping between sparsely sampled MR images and their fully sampled correspondences. In this study, with Institutional Review Board approval and HIPAA compliance, three hundred and sixty three-dimensional datasets of cartilage MRI

were acquired using UTE MRI and retrospectively undersampled in different acquisition schemes. During the training of the neural network, the optimal parameters were determined by iteratively minimizing the discrepancy between the predicted images and the fully sampled ground truth images via an advanced gradient descent method. After the neural network was trained, high quality images were predicted from sparsely sampled test images, which subsequently passed through the data consistency enforcement to form the final reconstructed images. The workflow is illustrated in Figure 1.

### Image Acquisition and Retrospective Undersampling

Three hundred and sixty volumetric cartilage MR images were obtained at the University of California San Diego, each consisting of 32 slices [30]. The images were acquired on a 3 Tesla scanner (GE Healthcare, Waukesha, WI) using an adiabatic inversion recovery spin-lock prepared UTE sequence with different numbers of IR spin-lock pulses (2, 4, 6, 8, 12, and 16). Other imaging parameters were as follows: echo time of 32  $\mu$ s, repetition time of 500 ms, flip angle of 10°, in-plane resolution of 256  $\times$  256, and voxel size of 0.586  $\times$  0.586  $\times$  3 mm<sup>3</sup>.

Given fully sampled images, a pseudo-random variable-density Cartesian acquisition, CIRCULAR Cartesian UnderSampling (CIRCUS) [31], was simulated as illustrated in Figure 2(a). Sparse sampling was performed on the ky-kz plane with partial acquisition (75%) applied in both directions. An acceleration factor of 4, 6 or 8 was achieved. The sparsely sampled k-space data were zero filled (without density compensation applied) and transformed back to the image domain using a three-dimensional inverse Fourier transform.

Similarly, the stack-of-stars radial acquisition [32] was simulated as shown in Figure 2(b). For each slice, the Radon transform was taken at undersampled projection angles, achieving an acceleration factor of 4, 6 or 8. The profiles were back-projected using the inverse Radon transform, and used as the input to the neural network.

### Reconstruction Using Convolutional Neural Network

A deep convolutional neural network framework was proposed for the image reconstruction of sparsely sampled MRI. In general, deep neural networks could provide a mapping that enforced data consistency and image sparsity in a joint optimization, where predicted images iteratively approached fully sampled images  $x$ , as described by

$$\min_{\theta} \|x - f_{cnn}(x_{ZF}|\theta)\|$$

Here,  $f_{cnn}$  was the forward mapping of the convolutional neural network that took the zero filled image  $x_{ZF}$  as input.  $f_{cnn}$  was parameterized by  $\theta = \{W_i, B_i, \alpha_i\}$ , where  $W_i$  and  $B_i$  represented the weights and biases in convolution operations, as defined in  $F_i(y) = W_i * F_{i-1}(y) + B_i$ , and  $\alpha$  was the coefficient defined in the PReLU (Parametric Rectified Linear Unit) function for nonlinear activation,  $PReLU = \max(x, 0) + \alpha \min(0, x)$  [33].

Throughout the network, volumetric processing was employed to exploit spatial continuity in three-dimensional space. This particularly facilitated acceleration in the through-plane

direction. Moreover, volumetric processing was applied to the whole image rather than to individual patches, which not only improved the computation efficiency by avoiding redundant computation between adjacent patches, but also provided valuable contextual information for signal synthesis and global artifact removal.

### Architecture of the Neural Network

To reconstruct high-quality images from their sparsely sampled correspondence, a hierarchical convolutional neural network was developed, as shown in Figure 3. After the sparsely sampled input images were fed into the network, image features were extracted and reorganized at multiple levels with different resolutions. In addition, global shortcut connections were established between the corresponding levels of the two paths, whereas local shortcut connections were established within the same level of a single path.

The hierarchical network had eleven levels. At each level, the resolution of feature maps was kept the same, and there were three convolutional blocks. Each convolutional block was composed of a convolutional layer and a nonlinear activation layer. At the convolutional layer, image features were extracted using  $3\times 3\times 3$  convolutional kernels, followed by zero-padding that kept the size of the feature map constant. At the nonlinear activation layer, the PReLU (Parametric Rectified Linear Unit) function was applied. All the data passed through two paths - a contracting path on the left and a subsequent expanding path on the right. At the next level along the contracting path, the resolution of feature maps was halved, and the number of filters was doubled. Down-sampling was accomplished using  $2\times 2\times 2$  convolutional kernels with a stride of 2 to replace the conventional max-pooling function, as suggested by [34]. On the contrary, at the following level along the expanding path, the resolution of feature maps was doubled, and the number of filters was halved; up-sampling was accomplished using  $2\times 2\times 2$  convolutional kernels as well. Finally, a  $1\times 1\times 1$  convolution kernel was adopted to merge information from multiple feature maps into one output image. Throughout this series of processing, the receptive field was continuously increased with convolution operations, resulting in large context with global constraints.

Moreover, ‘global’ shortcut connections were established between the two paths, aimed to compensate for the detailed information lost in down-sampling [35]. Meanwhile, ‘local’ shortcut connections were established within the same level of a single path to facilitate residual learning. In fact, relatively dense local shortcut connections were constructed by forwarding the input of a hierarchical level to all the subsequent convolutional blocks at the same level, unlike U-net that had no local shortcut connections or V-Net that had local simple shortcut connections, as compared in Figure 4. For shortcut connections, pointwise addition was adopted, where nonlinear activation was applied before the addition and identity mapping was conducted after the addition, as shown in Figure 5. This pre-addition activation scheme provided faster error reduction and lower training loss than the post-addition activation method [36].

### Data Consistency Enforcement

While the network output provided reasonable estimates for k-space coefficients at all data points, it was more accurate to replace the predictions by the original measurements at the

data points that were actually sampled. This enforcement of data consistency was incorporated into the loss function, which was the Root Mean Squared Error (RMSE):  $RMSE = \sqrt{\left(\sum_{i=1}^n \|z_i - x_i\|^2\right)}$ . Notice that  $x$  was the fully sampled image, and  $z$  was the predicted image with data consistency enforced, as formulated by  $z_i = IFT\{h(k_i, \hat{k}_i)\}$ . Here,  $k$  was the measured k-space data,  $\hat{k}$  was the predicted k-space data, which was the Fourier transform of the network prediction at the current iteration,  $\hat{k} = FT(f_{cnn}(x_{ZF}|\theta))$ , and  $h$  was the data consistency enforcement function defined as

$$h(k_i, \hat{k}_i) = I_{\{0\}}(k_i) * \hat{k}_i + (1 - I_{\{0\}}(k_i)) * k_i$$

where  $I$  was the indicator function. Alternatively, the data consistency enforcement function can be intuitively expressed as

$$h(k_i, \hat{k}_i) = \begin{cases} k_i & \text{if } k_i \neq 0 \\ \hat{k}_i & \text{if } k_i = 0 \end{cases}$$

### Training and Testing of the Neural Network

The neural network was trained to learn the optimal values of model parameters that were defined in convolutional filters or PReLU functions. The parameters in this deep neural network (total of 122,094) were initialized using the He method [33]. Errors between the reconstructed images  $z$  and the ground truth images  $x$  were back-propagated [16]. Parameters at all layers were updated accordingly using the Adam optimization method [37], which offered faster convergence than conventional stochastic gradient descent methods. We used an adaptive learning rate (starting from 0.001, halved every 2000 iterations),  $\beta_1$  of 0.9,  $\beta_2$  of 0.999, and  $\epsilon$  of  $10^{-8}$ .

Given the trained neural network, test images were reconstructed, where undersampled test images first passed through the neural network processing, and then experienced the subsequent data consistency enforcement. Additionally, compressed sensing reconstruction using the Split Bregman method [13, 38] was applied to the undersampled k-space data for comparison. The performance of the prediction was evaluated both qualitatively and quantitatively. A quantitative metric, structural similarity index measure (SSIM), was measured to assess image quality, which was defined as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where  $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ , and  $\sigma_y$  corresponded to the mean and standard deviation of signal intensity in the reconstructed image and the ground truth, whereas  $C_1$  and  $C_2$  were constants [37].

The network was implemented on a tensor-flow [39] based AI platform NiftyNet [40]. All computations were performed on a desktop computer running Linux operating system with an Intel i77700K CPU (4.2 GHz, and 32GB memory) and Nvidia GPU GeForce GTX1070.

## Results

The proposed T-net was trained with 336 three-dimensional images, each consisting of 32 slices. With the support of all the advanced deep learning techniques employed, this large-scaled neural network had fast convergence with a rapidly decreasing RMSE. After the network was trained, 24 three-dimensional images were tested. When the k-space data was reduced to one fourth of the fully sampled case, the images reconstructed using T-Net demonstrated high fidelity with the ground truth. Figure 6 showed the CIRCUS pseudo-random Cartesian sampling case, and Figure 7 demonstrated the stack-of-stars radial sampling case. From left to right, the three columns corresponded to the zero filled (input), fully sampled (ground truth), T-Net reconstructed, and compressed sensing images, respectively. Each row represented an individual subject. The T-Net reconstructed images had significantly improved image quality as compared to the zero filled images and compressed sensing images - the micro-structures, textures, and edges were substantially recovered with an improved SNR and suppressed artifacts.

As the acceleration factor was increased from 4 to 6 and 8, the quality of T-Net reconstructed images was slightly degraded. However, substantial details were recovered from the very blurry zero filled images. Compared with the compressed sensing images, the T-Net reconstructed images had more high frequency details, improved SNR and suppressed artifacts. Figure 8 compared the T-net reconstructed images with zero filled and compressed sensing images, when different acceleration factors were achieved via retrospective undersampling in a pseudorandom Cartesian sampling pattern (CIRCUS). Figure 9 demonstrated the case in which various acceleration factors were obtained using the stack-of-stars radial sampling.

The high degree of image fidelity was further confirmed by quantitative measurement of SSIM, as shown in Figure 10. With a given acceleration factor of 4, the SSIM of images reconstructed using T-Net was higher than the SSIM of zero filled images and compressed sensing images in both CIRCUS Cartesian acquisition and stack-of-stars radial acquisition, demonstrating the efficacy of the proposed method. The SSIM of the radial acquisition was slightly higher than that of the Cartesian acquisition.

The influences of shortcut connections were investigated as well. Figure 11 compared the images reconstructed using hierarchical deep neural networks with different shortcut connection patterns, when retrospective Cartesian undersampling was applied with an acceleration factor of 4. The image reconstructed with relatively dense local shortcuts (as proposed in T-Net) was the most similar to the fully sampled ground truth images, better than the one reconstructed without local shortcuts (as in U-Net) or the one reconstructed with simple local shortcuts (as in V-Net). The same trend was confirmed quantitatively by measuring the average SSIMs obtained using different approaches, as shown in Table 1. The highest SSIM was achieved in images reconstructed with relatively dense local shortcuts (as employed in T-Net) than those obtained without local shortcuts (as in U-Net) or with simple shortcuts (as in V-Net).

The effect of data consistency enforcement was demonstrated in Figure 12. The image obtained with data consistency enhancement had an apparently improved image quality as compared to the one without it in terms of reduced artifacts and better recovered micro-structures.

After the network was trained, reconstruction for a three-dimensional image took as short as 0.3 second.

## Discussion

In this study, we proposed a deep learning based reconstruction strategy to establish the mapping between undersampled images and high-quality MR images. Instead of taking a transform based on standard basis functions, deep learning based processing was data driven, taking advantage of *a priori* information learned from a large population. The deep learning based image reconstruction method proposed here was an end-to-end joint optimization that promoted the sparsity of feature maps and enforced data consistency. It was observed that sparsity was explicitly imposed by convolutional layers [28]. Particularly, in the proposed hierarchical neural network, feature maps were extracted at different scales, resulting in higher degree of sparsity (similar to wavelet processing) and potentially improving the maximal degree of undersampling [1]. The transform for sparse representation was learned rather than explicitly enforced. Therefore, the end-to-end optimization should provide a better solution than the frameworks that only incorporated deep learned *a priori* information.

A novel deep neural network was proposed and named as T-Net, since it was designed for MRI reconstruction applicable to  $T_1$  or  $T_2$  weighted images. Throughout the T-Net, volumetric processing was employed to exploit spatial continuity in three-dimensional space and support acceleration in the through-plane direction. Volumetric processing had been demonstrated to outperform two-dimensional processing with the same network architecture [35]. Moreover, the volumetric processing was applied to the whole image (rather than to individual patches as conducted in many super-resolution studies), providing valuable contextual information for signal synthesis and global artifact removal. In T-Net, fully connected network layers were not used. Converting a 2D/3D volume into 1D vector not only led to loss of spatial structures, but also increased the number of model parameters, which would consequently demand much more training data and more GPU memory. Furthermore, fully connected layers did not impose sparsity as convolutional layers [28].

T-Net has a hierarchical network architecture with global and local shortcut connections established for residual learning. Residual learning [41] has been shown for improving the performance of neural networks in a variety of tasks, including but not limited to image recognition [41], segmentation [17, 35, 42] and super-resolution [43]. Intuitively, it is easier to learn a residual image than the corresponding feature map since the former is much sparser [41]. In T-Net, relatively dense local shortcut connections were employed, going beyond the influential U-Net [17] and V-Net [44]. This was first inspired by the dense shortcut connections employed in Dense Net [45], which forwarded the output of every convolutional block to all the subsequent blocks. However, the necessity of large amount of



memory was problematic for 3D high resolution image sets. Motivated by an alternative shortcut connection pattern proposed in Deep Recursive Residual Network [43], we established ‘relatively dense’ local shortcut connections by forwarding the input of a hierarchical level to all the subsequent convolutional blocks at the same hierarchical level, reaching a good balance between the network performance and memory consumption. The comparison of different local short connection patterns is illustrated in Figure 13. In the context of image reconstruction, another interesting finding was that the local shortcuts at the first level of the contracting path slightly degraded the quality of reconstructed images, which may be due to the fact that the undersampling artifacts and blurring observed in input images were propagated to subsequent feature maps. Hence, the local shortcuts at the first level of the contracting path were excluded. The relatively dense local shortcut connections worked effectively, as demonstrated in Figure 11 and Table 1.

In this study, the T-Net was trained with MRI images that were consistently acquired. Domain adaptation, domain transfer, or more generically called transfer learning [16], was commonly applied in medical imaging due to limited data available for network training. Those neural networks were pre-trained using natural images or images in other sensor domain and fine-tuned with images in the same sensor domain. Although domain adaptation was effective in general, a neural network precisely trained with images in the same sensor domain would be more powerful since the model was adapted to input images that had realistic patterns of artifacts and noise. We collected large quantity of MRI images that were consistently acquired, and simulated undersampled images by obtaining k-space data from dicom images, which were not totally realistic. However due to the limited availability of k-space data that were consistently acquired, the simulation still provides valuable insights on the degree of acceleration that could be possibly achieved as well as the influence of different sampling patterns and generalizable to real MRI experiments. While the network was trained using only healthy volunteers, we believe that it will not fail to reconstruct or change abnormal anatomic structure that did not exist in the training data.

We explored the flexibility of sampling trajectories, which was specific to MRI as compared to other imaging modalities. Two undersampling schemes were simulated – CIRCUS and stack-of-stars. CIRCUS mainly suffered from signal loss in some micro-structures, whereas stack-of-stars experienced undersampling streak artifacts. By visual inspection, removing streak artifacts in radial acquisition seemed to be more challenging. But in quantitative measurement, the stack-of-stars radial acquisition had a higher average SSIM than the pseudo-random Cartesian sampling. We started from zero filled images without conducting density compensation, which was not ideal for achieving a high acceleration factor but helped us to concentrate on the proposed deep learning strategy without considering the influence of other incorporated techniques. In both radial and cartesian acquisitions, an acceleration factor of 8 was achieved in cartilage UTE MRI. Higher acceleration factors could be possible using alternative sampling patterns or larger training sets, or combing with parallel imaging.

The proposed framework is a generic MRI reconstruction approach that can be applied to other pulse sequences and benefit various clinical applications. The reduction in scan time will be highly appreciated especially in time consuming scans, such as dynamic MRI that

requires high temporal resolution and quantitative MRI that are currently clinically infeasible due to their extended scan times. In addition, instantaneous reconstruction is another major advantage of the proposed method. While the training of the neural network takes time, the reconstruction of test images can be completed in real time, which facilitates its utility in clinical practice.

The current approach can be extended in several directions. First, a neural network can provide a direct mapping from k-space to the image domain. In addition, other loss functions can be adopted, such as the normalized MSE, L1 norm, SSIM, and mutual information. These efforts can potentially further improve the quality of reconstructed images.

## Conclusions

The reported study established a generic deep learning framework for MR image reconstruction, which could significantly accelerate MR image acquisition. A volumetric hierarchical deep residual convolutional neural network, T-Net, was constructed to provide an end-to-end mapping from sparsely sampled images to high quality output images in real time. In cartilage MRI acquired using UTE and retrospectively undersampled using various sampling schemes, an acceleration factor of 8 was achieved, where reconstructed images had high fidelity to the ground truth with limited artifacts and high SNR. The proposed deep learning based image reconstruction method has the potential to be extended to a variety of MRI acquisition techniques as well as other imaging modalities (e.g. CT or PET).

## Acknowledgements

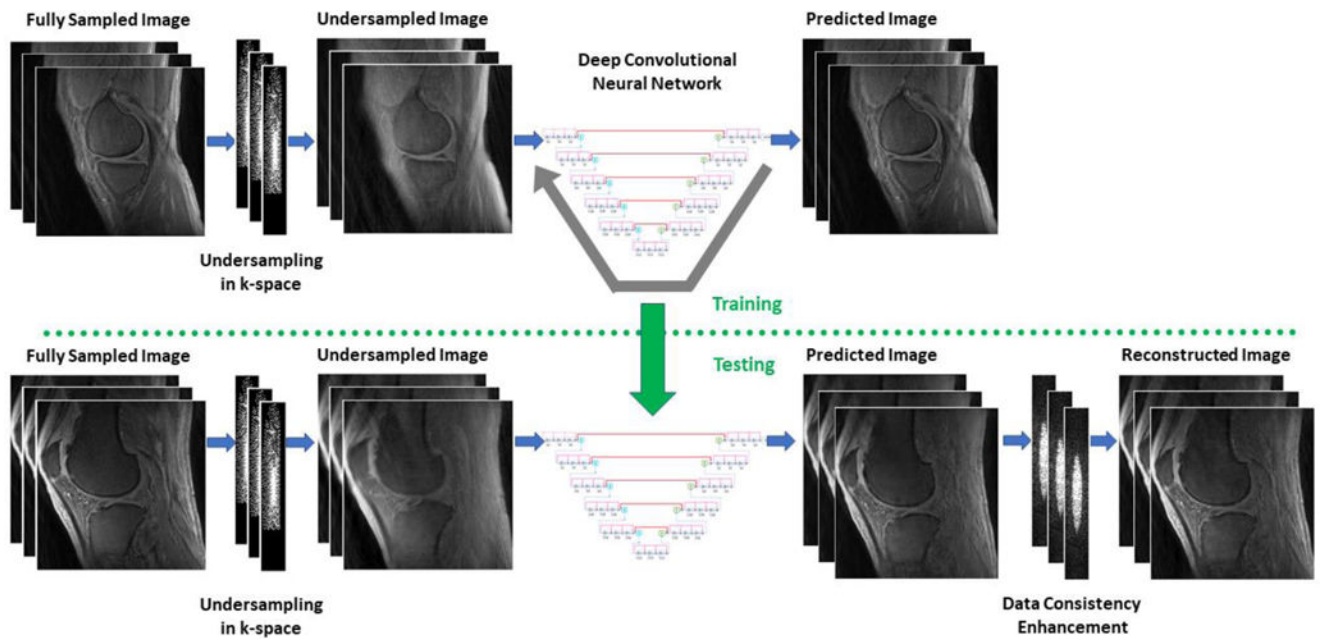
The author would like to thank Drs. Charles Mistretta, Marian Chan, Tianliang Gu, and Annie Hsu for their helpful discussion. The research was supported by NIH/NCI (1R01 CA176553) and Faculty Research Award from Google Inc.

## Reference

- [1]. Lustig M, Donoho D, and Pauly JM, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic resonance in medicine*, vol. 58, pp. 1182–1195, 2007. [PubMed: 17969013]
- [2]. Qu X, Cao X, Guo D, Hu C, and Chen Z, "Combined sparsifying transforms for compressed sensing MRI," *Electronics letters*, vol. 46, pp. 121–123, 2010.
- [3]. Lai Z, Qu X, Liu Y, Guo D, Ye J, Zhan Z, et al., "Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform," *Medical image analysis*, vol. 27, pp. 93–104, 2016. [PubMed: 26096982]
- [4]. Ravishankar S and Bresler Y, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE transactions on medical imaging*, vol. 30, p. 1028, 2011. [PubMed: 21047708]
- [5]. Huang Y, Paisley J, Lin Q, Ding X, Fu X, and Zhang X-P, "Bayesian nonparametric dictionary learning for compressed sensing MRI," *IEEE Transactions on Image Processing*, vol. 23, pp. 5007–5019, 2014. [PubMed: 25265609]
- [6]. Feng L, Grimm R, Block KT, Chandarana H, Kim S, Xu J, et al., "Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magnetic resonance in medicine*, vol. 72, pp. 707–717, 2014. [PubMed: 24142845]

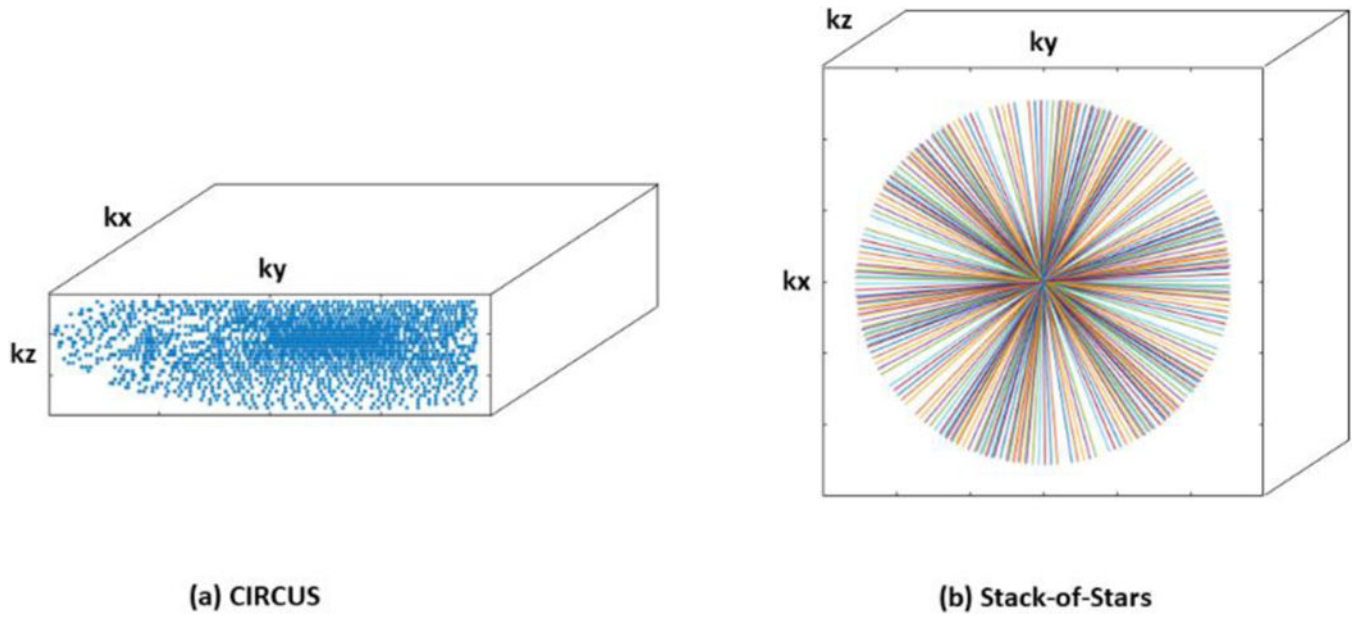
- [7]. Otazo R, Kim D, Axel L, and Sodickson DK, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magnetic resonance in medicine*, vol. 64, pp. 767–776, 2010. [PubMed: 20535813]
- [8]. Sharma SD, Fong CL, Tzung BS, Law M, and Nayak KS, "Clinical image quality assessment of accelerated magnetic resonance neuroimaging using compressed sensing," *Investigative radiology*, vol. 48, pp. 638–645, 2013. [PubMed: 23538890]
- [9]. Hollingsworth KG, Higgins DM, McCallum M, Ward L, Coombs A, and Straub V, "Investigating the quantitative fidelity of prospectively undersampled chemical shift imaging in muscular dystrophy with compressed sensing and parallel imaging reconstruction," *Magnetic resonance in medicine*, vol. 72, pp. 1610–1619, 2014. [PubMed: 24347306]
- [10]. Jaspan ON, Fleysher R, and Lipton ML, "Compressed sensing MRI: a review of the clinical literature," *The British journal of radiology*, vol. 88, p. 20150487, 2011.
- [11]. Vaswani N and Lu W, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4595–4607, 2010.
- [12]. Lu W and Vaswani N, "Modified compressive sensing for real-time dynamic MR imaging," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 3045–3048.
- [13]. Gopi VP, Palanisamy P, Wahid KA, and Babyn P, "MR image reconstruction based on iterative split Bregman algorithm and nonlocal total variation," *Computational and mathematical methods in medicine*, vol. 2013, 2013.
- [14]. Jung H, Sung K, Nayak KS, Kim EY, and Ye JC, "k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI," *Magnetic resonance in medicine*, vol. 61, pp. 103–116, 2009. [PubMed: 19097216]
- [15]. Mistretta CA, Wieben O, Velikina J, Block W, Perry J, Wu Y, et al., "Highly constrained backprojection for time-resolved MRI," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 55, pp. 30–40, 2006.
- [16]. LeCun Y, Bengio Y, and Hinton G, "Deep learning," *nature*, vol. 521, p. 436, 2011.
- [17]. Ronneberger O, Fischer P, and Brox T, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [18]. Dong C, Loy CC, He K, and Tang X, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, pp. 295–307, 2016. [PubMed: 26761735]
- [19]. Xing L, Krupinski EA, and Cai J, "Artificial intelligence will soon change the landscape of medical physics research and practice," *Medical physics*, 2018.
- [20]. Qin W, Wu J, Han F, Yuan Y, Zhao W, Ibragimov B, et al., "Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation," *Physics in Medicine & Biology*, vol. 63, p. 095017, 2011.
- [21]. Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, et al., "Accelerating magnetic resonance imaging via deep learning," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, 2016, pp. 514–517.
- [22]. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, et al., "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic resonance in medicine*, vol. 79, pp. 3055–3071, 2018. [PubMed: 29115689]
- [23]. Yang Y, Sun J, Li H, and Xu Z, "ADMM-Net: A deep learning approach for compressive sensing MRI," *arXiv preprint arXiv:1705.06869*, 2017.
- [24]. Han YS, Yoo J, and Ye JC, "Deep learning with domain adaptation for accelerated projection reconstruction MR," *arXiv preprint arXiv:1703.01135*, 2017.
- [25]. Dar SUH and Çukur T, "A Transfer-Learning Approach for Accelerated MRI using Deep Neural Networks," *arXiv preprint arXiv:1710.02615*, 2017.
- [26]. Schlemper J, Caballero J, Hajnal JV, Price A, and Rueckert D, "A deep cascade of convolutional neural networks for MR image reconstruction," in *International Conference on Information Processing in Medical Imaging*, 2017, pp. 647–658.

- [27]. Mardani M, Gong E, Cheng JY, Vasanawala S, Zaharchuk G, Alley M, et al., “Deep generative adversarial networks for compressed sensing automates MRI,” arXiv preprint arXiv:1706.00051, 2017.
- [28]. Zhu B, Liu JZ, Cauley SF, Rosen BR, and Rosen MS, “Image reconstruction by domain-transform manifold learning,” *Nature*, vol. 555, p. 487, 2011.
- [29]. Lyu Q and Wang G, “Quantitative MRI: Absolute T1, T2 and Proton Density Parameters from Deep Learning,” arXiv preprint arXiv:1806.07453, 2018.
- [30]. Ma YJ, Carl M, Searleman A, Lu X, Chang EY, and Du J, “3D adiabatic T1ρ prepared ultrashort echo time cones sequence for whole knee imaging,” *Magnetic resonance in medicine*, 2018.
- [31]. Liu J and Saloner D, “Accelerated MRI with CIRCular Cartesian UnderSampling (CIRCUS): a variable density Cartesian sampling strategy for compressed sensing and parallel imaging,” *Quantitative imaging in medicine and surgery*, vol. 4, p. 57, 2011.
- [32]. Peters DC, Korosec FR, Grist TM, Block WF, Holden JE, Vigen KK, et al., “Undersampled projection reconstruction applied to MR angiography,” *Magnetic Resonance in Medicine*, vol. 43, pp. 91–101, 2000. [PubMed: 10642735]
- [33]. He K, Zhang X, Ren S, and Sun J, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
- [34]. Springenberg JT, Dosovitskiy A, Brox T, and Riedmiller M, “Striving for simplicity: The all convolutional net,” arXiv preprint arXiv:1412.6806, 2014.
- [35]. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, and Ronneberger O, “3D U-Net: learning dense volumetric segmentation from sparse annotation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 424–432.
- [36]. He KM, Zhang XY, Ren SQ, and Sun J, “Identity Mappings in Deep Residual Networks,” *Computer Vision - Eccv 2016, Pt Iv*, vol. 9908, pp. 630–645, 2016.
- [37]. Kingma DP and Ba J, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.
- [38]. Goldstein T and Osher S, “The split Bregman method for L1-regularized problems,” *SIAM journal on imaging sciences*, vol. 2, pp. 323–343, 2009.
- [39]. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al., “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” arXiv preprint arXiv:160304467, 2016.
- [40]. Gibson E, Li W, Sudre C, Fidon L, Shakir D, Wang G, et al., “NiftyNet: a deep-learning platform for medical imaging,” arXiv preprint arXiv:1709.03485, 2017.
- [41]. He K, Zhang X, Ren S, and Sun J, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [42]. Milletari F, Navab N, and Ahmadi S-A, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *3D Vision (3DV), 2016 Fourth International Conference on*, 2016, pp. 565–571.
- [43]. Tai Y, Yang J, and Liu X, “Image super-resolution via deep recursive residual network,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [44]. Milletari F, Navab N, and Ahmadi SA, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” *Proceedings of 2016 Fourth International Conference on 3d Vision (3dv)*, pp. 565–571, 2016.
- [45]. Huang G, Liu Z, Weinberger KQ, and van der Maaten L, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, p. 3.



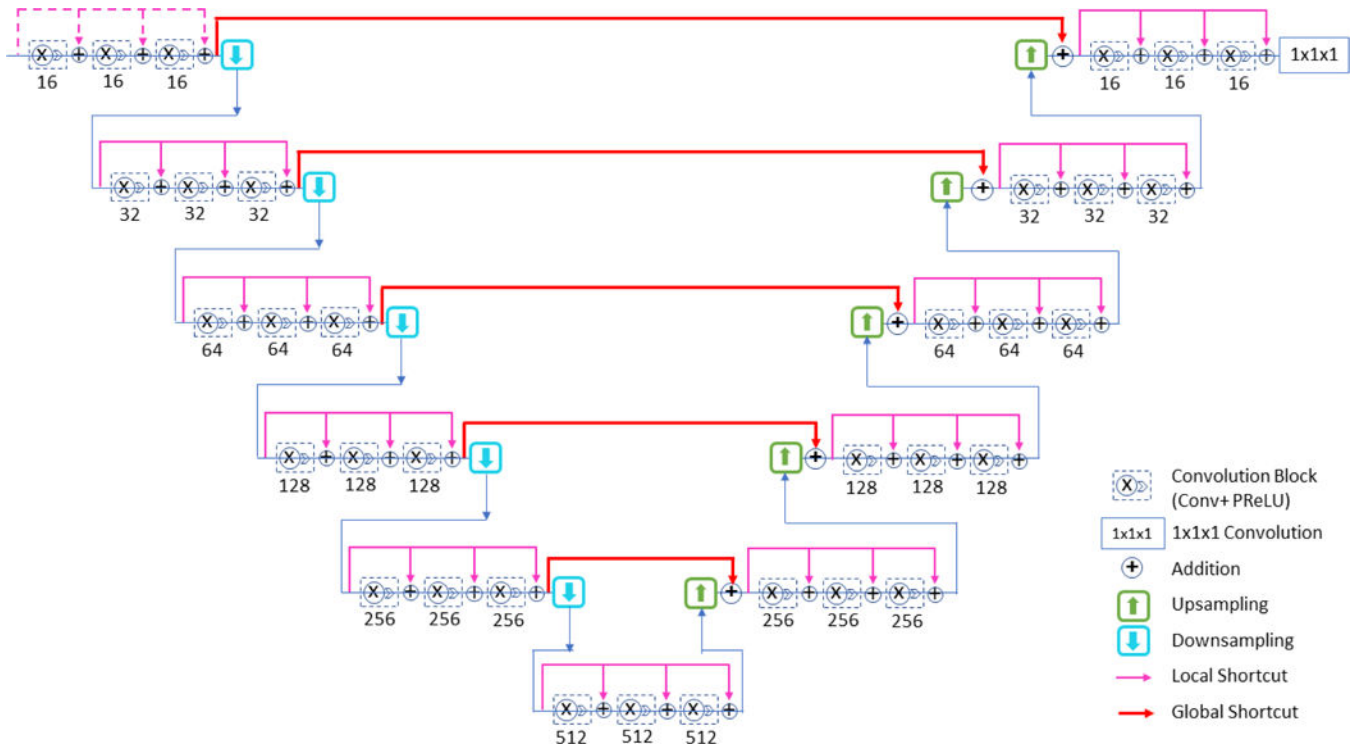
**Figure 1.**

The workflow of deep learning based MR reconstruction. MR images were retrospectively undersampled in k-space and transformed back to the image domain. A deep convolutional neural network was trained to provide a mapping from sparsely sampled zero-filled images to fully sampled high quality images, where the loss between predicted images and ground truth images was back-propagated and used to update model parameters. The trained network model was employed to predict high quality images from undersampled test images, whose output subsequently passed through data consistency enforcement to form the final reconstructed images.



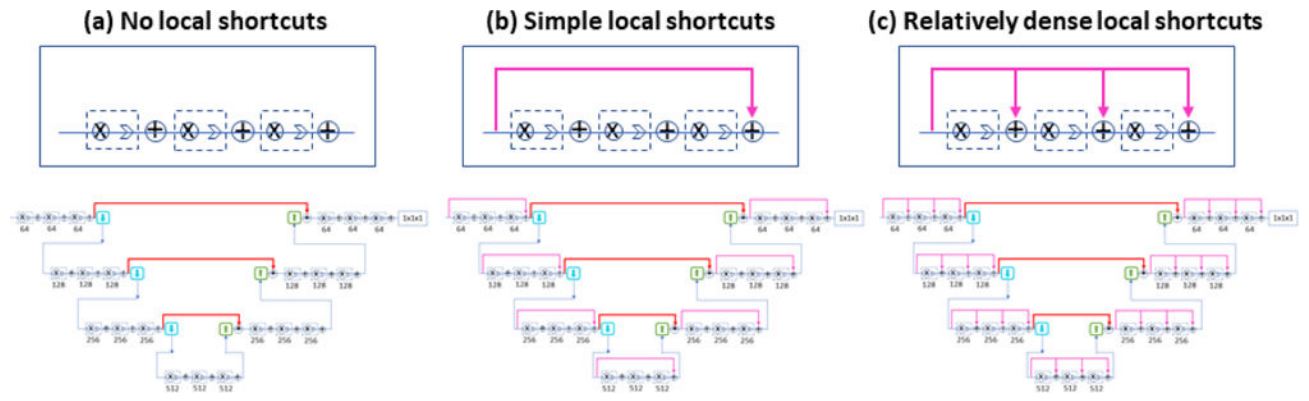
**Figure 2.**

Two k-space undersampling patterns retrospectively applied in this study. (a) CIRCUS: a pseudo-random variable-density Cartesian sampling pattern, where sparse sampling was performed on the ky-kz plane. (b) stack-of-stars radial sampling, where undersampling was conducted on the kx-ky plane.



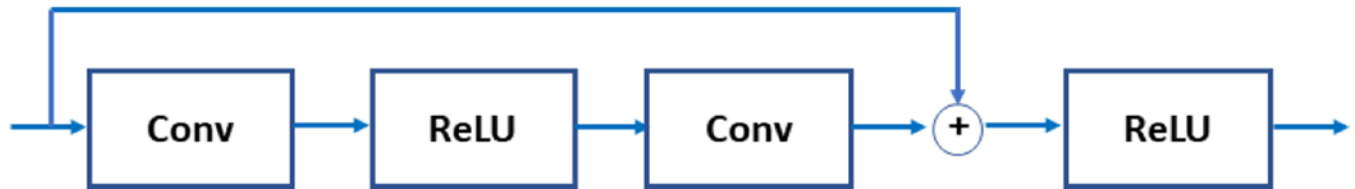
**Figure 3.**

The hierarchical architecture of the proposed deep convolutional neural network, T-Net. It was composed of a contracting path (on the left) and a subsequent expanding path (on the right). Along the contracting path, the resolution of feature maps shrank at the next level, and the number of feature maps or convolutional kernels doubled as indicated. Along the expanding path, the resolution of feature maps expanded at the subsequent level, and the number of feature maps halved as indicated. In this way, image features were extracted and reorganized at multiple levels with different resolutions. Global shortcut connections were established between the corresponding levels of the two paths, whereas local shortcut connections were constructed within the same level of a single path. Finally, after convolving with a  $1 \times 1 \times 1$  kernel, output MR images were predicted.

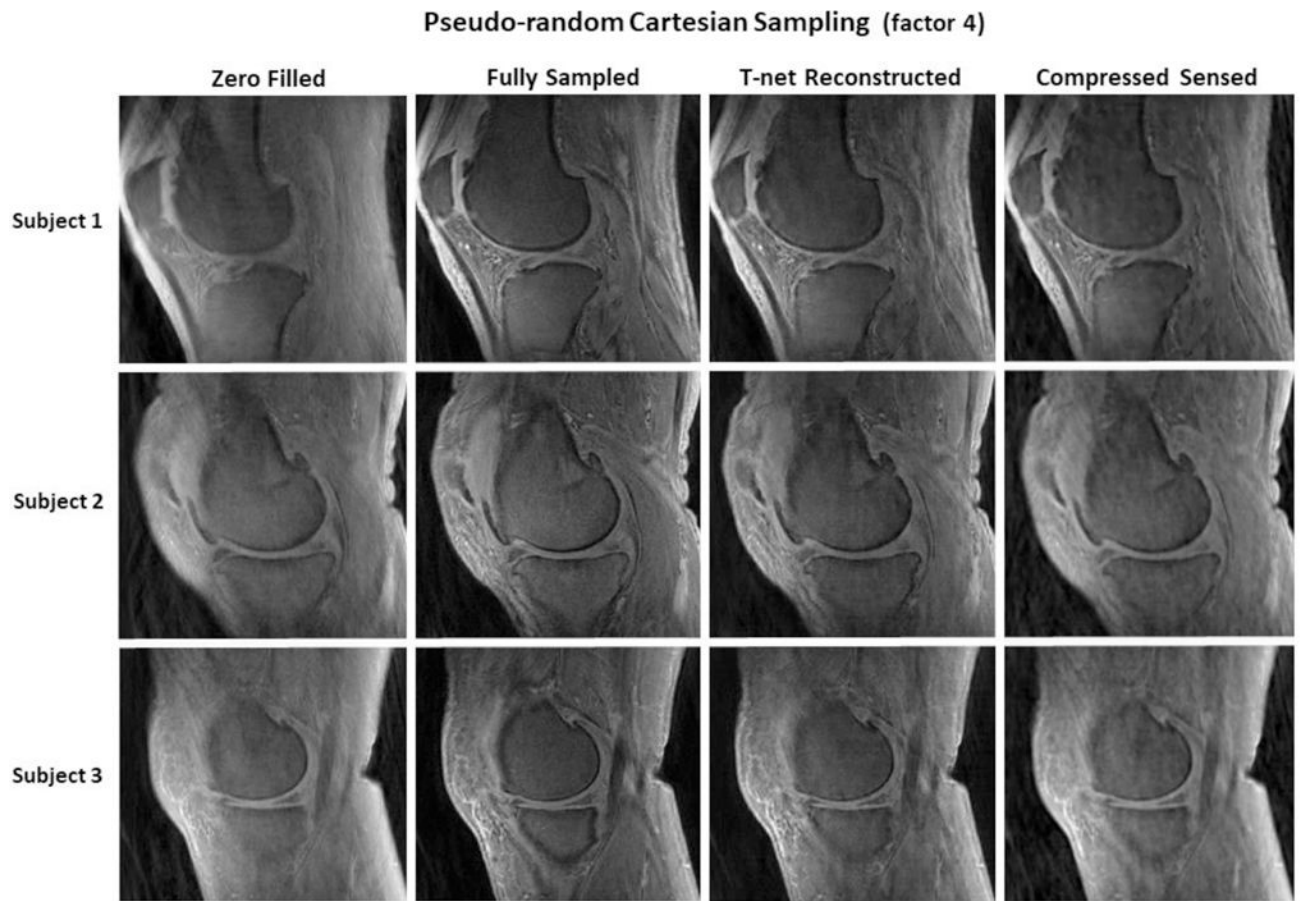


**Figure 4.** Comparison of the proposed local shortcut connection scheme in T-Net with the ones adopted in U-Net and V-Net. (a) no local shortcuts, as in U-Net (b) simple local shortcuts (forwarding the input of a hierarchical level to the output at the same level), as in V-Net (c) relatively dense local shortcuts (forwarding the input of a hierarchical level to all the subsequent convolutional blocks at the same level), as proposed in T-Net.



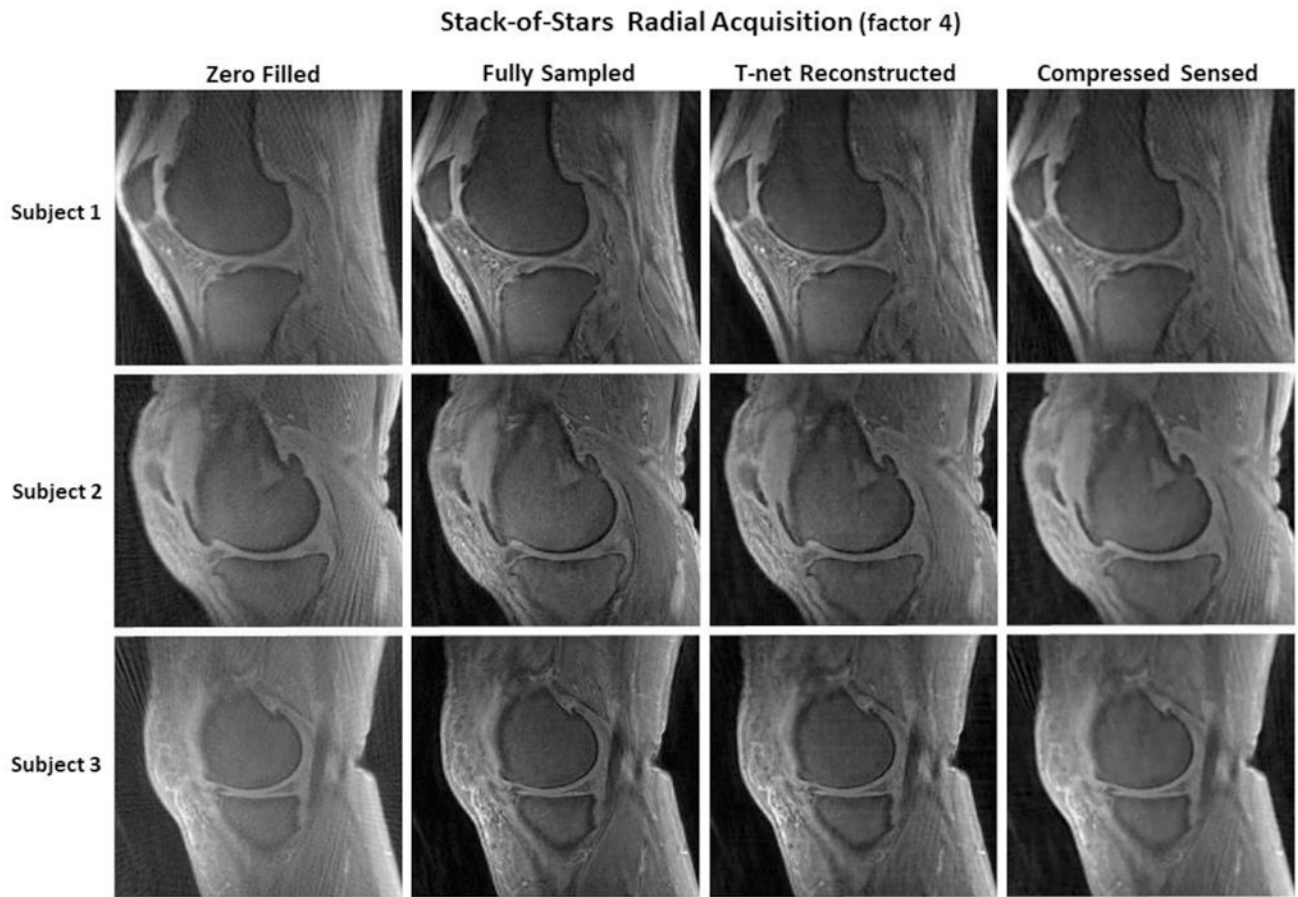
**(a) Pre-addition Activation****(b) Post-addition Activation**

**Figure 5.** Comparison of pre-addition activation and post-addition activation in residual learning. The pre-addition activation scheme was adopted in T-Net with identity mapping applied after addition, which was reported to provide faster error reduction and lower training loss than the conventional post-addition activation scheme.



**Figure 6.**

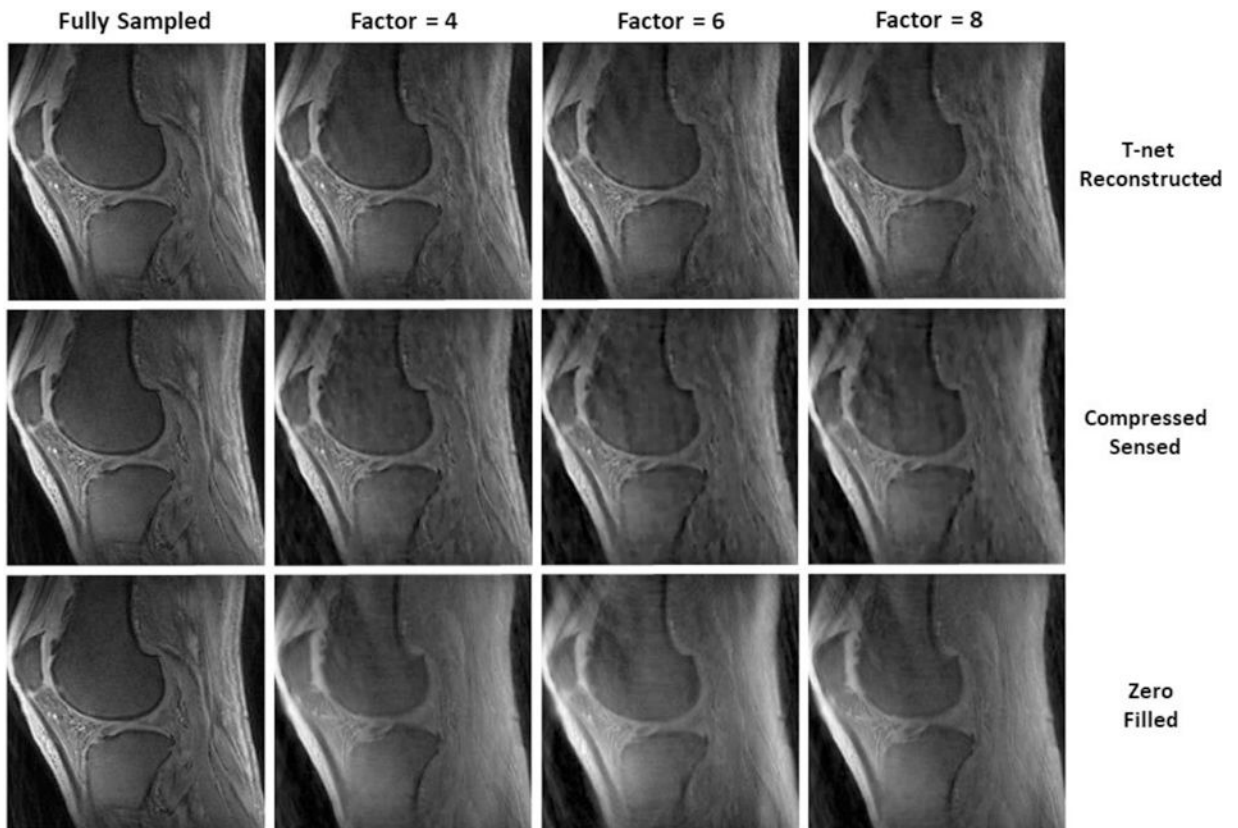
Comparison of zero filled, fully sampled, T-net reconstructed, and compressed sensing images, which were retrospectively undersampled in the CIRCUS pseudo-random Cartesian acquisition scheme with an acceleration factor of 4 achieved. Each row represented an individual subject. The micro-structures lost in the zero-filled images were significantly recovered in the T-net reconstructed images, which had high fidelity with the ground truth. Furthermore, the image quality of the T-Net reconstructed images was improved as compared to that of the compressed sensing images.



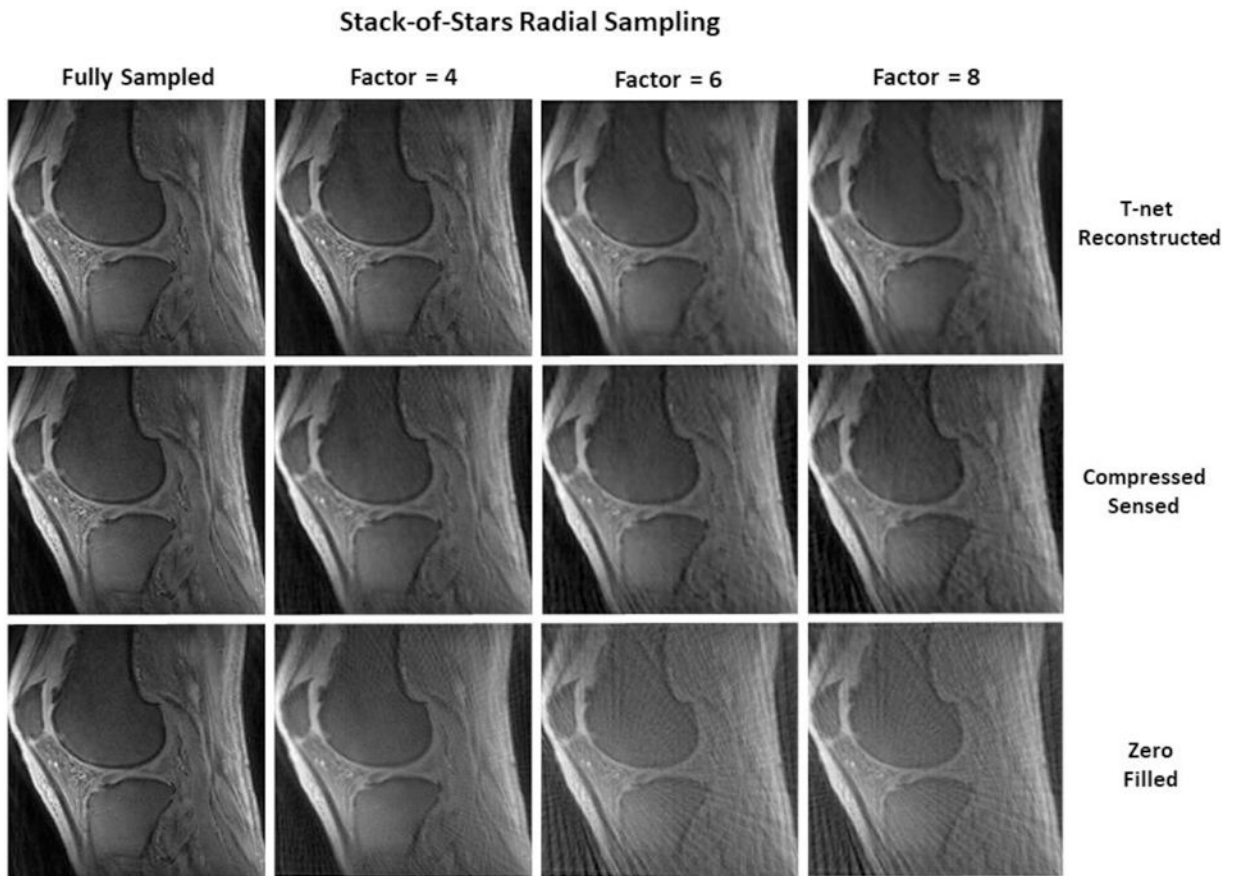
**Figure 7.**

Comparison of zero filled, fully sampled, T-net reconstructed, and compressed sensing images, which were retrospectively undersampled in the stack-of-stars radial acquisition scheme with an acceleration factor of 4 achieved. Each row represented an individual subject. The undersampling streak artifacts appearing in the zero filled images were significantly suppressed in the T-Net reconstructed images, which was more consistent with the ground truth. Additionally, the image quality of the T-Net reconstructed images was improved as compared to that of the compressed sensing images.

## Pseudo-random Cartesian Sampling

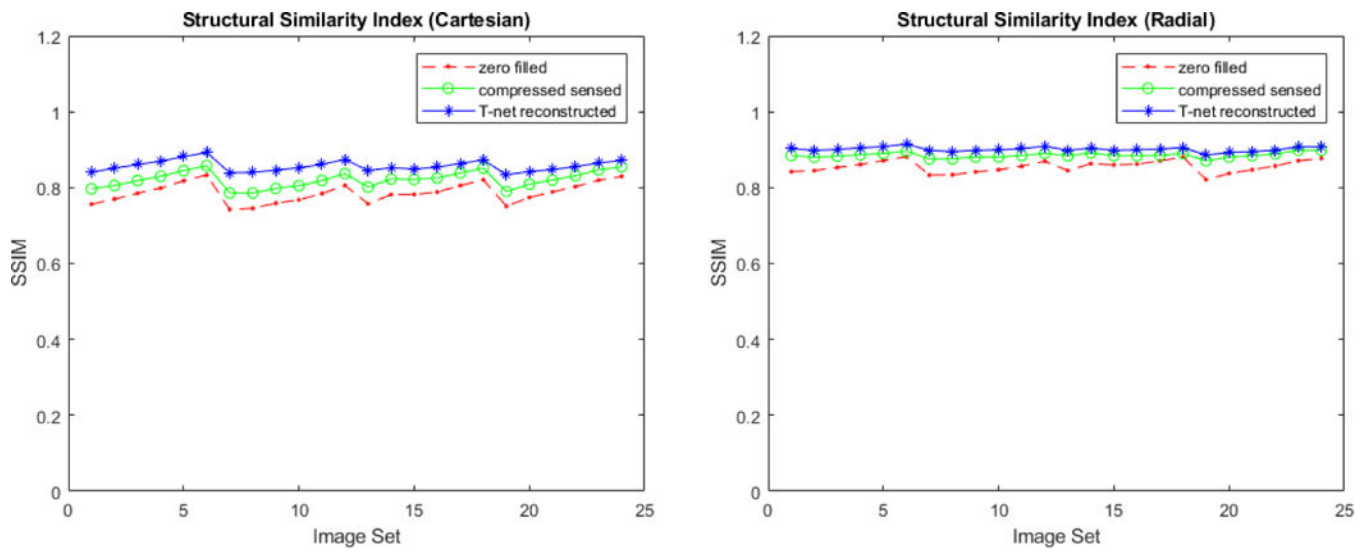
**Figure 8.**

Comparison of T-net reconstructed, compressed sensing, and zero filled images, which were obtained via retrospective undersampling in the CIRCUS pseudo-random Cartesian acquisition pattern with different acceleration factors achieved. As the acceleration factor was increased from 4 to 6 and 8, the quality of T-Net reconstructed images was slightly degraded with some micro-structures hard to differentiate (in the posterior regions of the knee). However, substantial details were recovered from the very blurry zero filled images. Even when compared with the compressed sensing images, the T-Net reconstructed images still had more high frequency details, improved SNR and suppressed artifacts.



**Figure 9.**

Comparison of T-net reconstructed, compressed sensing, and zero filled images, which were obtained via retrospective undersampling in the stack-of-stars radial acquisition pattern with different acceleration factors achieved. As the acceleration factor was increased from 4 to 6 and 8, global undersampling streak artifacts and local blurring (loss of micro-structures in the posterior regions of the knee) became more obvious across all images. In the T-Net reconstructed images, substantial high frequency details were recovered with SNR increased and streak artifacts suppressed. The T-Net reconstructed images had apparently improved image quality as compared to the compressed sensing images.



**Figure 10.**

The SSIM of images acquired with (a) the CIRCUS Cartesian sampling and (b) the stack-of-stars radial sampling. For a given acceleration factor of 4, the SSIM of T-Net reconstructed images was higher than that of zero filled images and compressed sensing images. The SSIM of the radial acquisition was slightly higher than that of the Cartesian acquisition.

## Images Reconstructed Using the Deep Neural Network with Various Local Shortcut Connections



**Figure 11.**

Comparison of images reconstructed using hierarchical deep neural networks with different shortcut patterns, when retrospective Cartesian undersampling was applied with an acceleration factor of 4. (a) ground truth, (b) image reconstructed without local shortcuts (as in U-Net), (c) image reconstructed with simple local shortcuts (as in V-Net), (d) image reconstructed with relatively dense local shortcuts (as in T-Net). The image reconstructed with relatively dense local shortcuts was the most similar to the ground truth.

**Images Reconstructed Using the Network with/without Data Consistency Enforcement****Figure 12.**

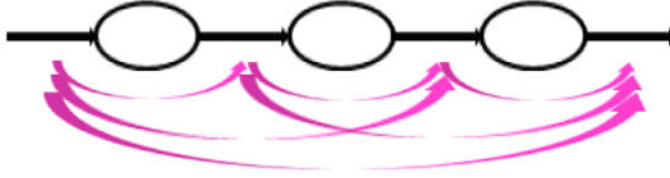
Comparison of images reconstructed using T-Net with or without data consistency enforcement. (a) ground truth, (b) images reconstructed without data consistency enforcement, and (c) images reconstructed with data consistency enhancement. The data consistency enforcement helped to improve the image quality. The undersampling artifacts were reduced, and the lost micro-structures were better recovered.



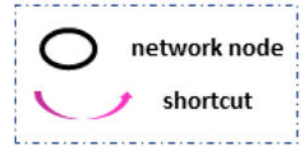
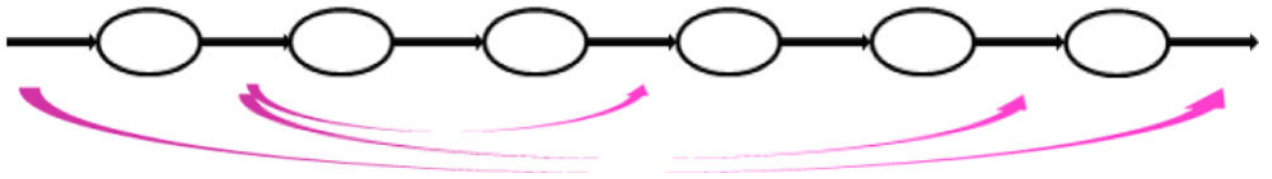
(a) T-Net



(b) Dense Net



(c) Deep Recursive Residual Network



**Figure 13.** Several dense shortcut connection schemes that motivated the design of T-Net. (a) T-Net, in which the input of a network level was forwarded to all the subsequent convolutional blocks at the same level, (b) Dense Net, in which the output of every convolutional block was forwarded to all the subsequent blocks, and (c) Deep Recursive Residual Network, which had shortcut connections with various ranges of influence. Here, the origins of the shortcuts were close to the input of the network level.

**Table 1.**

The average SSIMs of images reconstructed using deep neural networks with different shortcut connections. Images reconstructed with relatively dense local shortcut connections (as adopted in T-Net) had a higher SSIM value than the ones reconstructed with simple local shortcuts (as in V-Net) or those reconstructed without local shortcuts (as in U-Net). This quantitative result was consistent with the observation in Figure 11.

<b>Image</b>	<b>Avg SSIM</b>
T-Net reconstructed images without local shortcuts	0.8484
T-Net reconstructed images with single local shortcuts	0.8549
T-Net reconstructed images with relatively dense local shortcuts	0.8603