ACCELERATING MAGNETIC RESONANCE IMAGING VIA DEEP LEARNING

Shanshan Wang[1], Zhenghang Su[2],Leslie Ying[3],Xi Peng[1],Shun Zhu[1]
Feng Liang[4], *Dagan Feng[5]* and *Dong Liang[1]*

[1]Paul C. Lauterbur Research Center for Biomedical Imaging, SIAT, CAS, Shenzhen, P.R.China
[2]School of Information Technologies, Guangdong University of Technology, Guangzhou, P.R. China
[3]Department of Biomedical Engineering and Department of Electrical Engineering,
The State University of New York, Buffalo, New York 14260, USA
[4]Department of Industrial Engineering, NanKai University, Tianjin, P.R. China
[5]School of Information Technologies, University of Sydney, Sydney, NSW 2006, Australia

ABSTRACT

This paper proposes a deep learning approach for accelerating magnetic resonance imaging (MRI) using a large number of existing high quality MR images as the training datasets. An off-line convolutional neural network is designed and trained to identify the mapping relationship between the MR images obtained from zero-filled and fully-sampled k-space data. The network is not only capable of restoring fine structures and details but is also compatible with online constrained reconstruction methods. Experimental results on real MR data have shown encouraging performance of the proposed method for efficient and effective imaging.

Index Terms— Deep learning, magnetic resonance imaging, prior knowledge, convolutional neural network

1. INTRODUCTION

Magnetic resonance imaging (MRI) is an indispensable tool for medical diagnosis, disease staging and clinical research due to its strong capability in providing rich anatomical and functional information and its non-radiation and non-ionizing nature. However, most of advanced applications such as cardiovascular imaging, functional MRI, magnetic resonance spectroscopy and parameter mapping are not yet widely used in clinic due to the long scanning time of MRI [1, 2].

To accelerate MR scans, efforts are mainly in three directions 1) physics based fast imaging sequences, 2) hardware based parallel imaging and 3) signal processing based MR image reconstruction from reduced samples. The combination of these techniques have also shown their appearance in a great number of publications [2, 3, 4, 5, 6]. The first two categories and a few techniques of the third category (e.g. partial Fourier) have already been applied in commercial scanners as a routine protocol for shortening the total scan time [2, 3, 4, 5, 6]. Signal processing based methods, explores prior information on MR images and utilize them to regularize the reconstruction from undersampled K-space measurements with the advantage of no physical, physiological and hardware restrictions. Sparsity is one commonly used prior information due to the emergence of Compressed sensing (C-S) and there are also other priors being considered, such as low-rank [7], statistics distribution regularization [8], manifold fitting [9], generalized series (GS) model [10] and so on. The prior information used can be roughly categorized into adaptive and non-adaptive ones. For example, total variation and Wavelet transform, singular value decomposition (SVD) are non-adaptive ones [3, 5, 4]; dictionary learning and data-driven tight frames are adaptive [11, 12, 13, 14] . Generally, adaptive priors can capture more structures while non-adaptive ones are more computationally efficient. Nevertheless, despite all the successes achieved by the aforementioned methods, it is easy to discover that they only explore the prior information either directly from the image to be reconstructed or with very few reference images involved. Considering the similarity on the anatomic information of the same organ/tissue between different people and the enormous images acquired every day, it is straightforward to collect many reference images and learn an off-line prior model to aid online fast imaging.

Deep learning, a technique attempting to model high-level abstractions in data with multiple processing layers, has shown explosive popularity in recent years with the availability of powerful GPUs. Especially, convolutional neural network (CNN) has exhibited its significance in addressing large-scale vision tasks such as action recognition [15], image classification [16], super-resolution [17] and denoising [18]. CNNs have quite a few merits, such as the lack of dependence on prior-knowledge, no need to design hand-engineered fea-

Grant support : China NSFC 11301508, 81120108012, 81328013, 61471350, 71271122 the Natural Science Foundation of Guangdong 2015A0202140192015A030310314, 2015A030313740, the Basic Research Program of Shenzhen JCYJ20150630114942318, JCYJ20140610152828678, JCYJ20140610151856736 and the youth innovation project of SIAT under Y4G0071001Y3G0151001, US NIH R21EB020861 for Ying and ARC grants

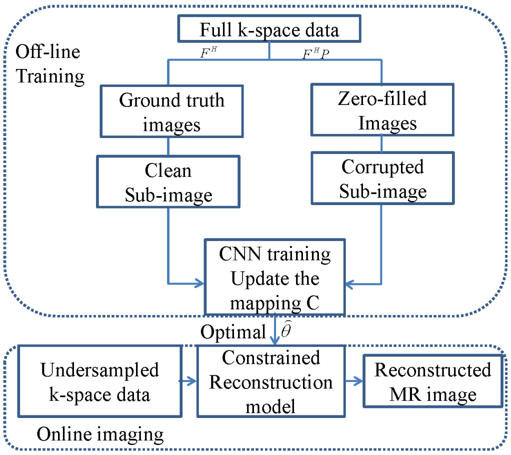

Fig. 1. The flowchart of the proposed method

tures and strong ability to capture image structures, which motivated us to employ it for MR image reconstruction from undersampled k-space data. What is more important is that the relationship between the zero-filled MR image and the ground-truth image can be interpreted in a convolutional way, which will be explained in the theory parts.

In this paper we propose an off-line convolutional neural network to learn an end-to-end mapping between zero-filled and fully-sampled MR images. This network is not only capable of restoring the details and fine structures of the MR images, but is also compatible with any online reconstruction algorithm for more efficient and effective imaging. We have tested the proposed method on a set of in-vivo MR data and the results have shown promising.

2. THEORY

This section provides the concept and theoretical foundation for learning the convolutional neural network for undersampled MR image reconstruction. Fig. 1 presents the flowchart of the proposed method.

2.1. Overall Training Formulation

Consider the undersampled raw K-space data as

$$f = PFu \tag{1}$$

where P is a diagonal matrix representing the undersampling mask, F denotes the full Fourier encoding matrix normalized as $F^H F = I$, u is the original (ground truth) image and therefore Fu represents the full k-space data. H represents the Hermitian transpose operation. The zero-filled MR image z is generated as the direct inverse transform of the observed data like

$$z = F^H PFu \tag{2}$$

As stated in [19], in terms of linear algebra, the circular convolution of a signal u with a pulse p can be written as $F^H PFu$, where P is a diagonal matrix whose non-zero entries are the Fourier transform of p.

We try to learn a fully convolutional neural network to restore accurate MR images from undersampled Fourier data. Given a pre-acquired dataset of MR corrupted/ground truth images, we try to minimize the following objective

$$\underset{\Theta}{\operatorname{argmin}} \left\{ \frac{1}{2\mathrm{T}} \sum_{t=1}^{T} \|C(z_t;\Theta) - u_t\|_2^2 \right\} \tag{3}$$

where C means the end-to-end mapping function with its hidden parameters $\Theta = \{(W_1, b_1), ..., (W_l, b_l), ..., (W_L, b_L)\}$ to be estimated and T is the total number of training samples. In order to increase the robustness of the network, we generate more training samples by separating the whole image pairs into overlapping subimage pairs $x_{t,n}$ and $y_{t,n}$ and minimize

$$\underset{\Theta}{\operatorname{argmin}} \left\{ \frac{1}{2\mathrm{TN}} \sum_{t=1}^{T} \sum_{n=1}^{N} \|C(x_{t,n};\Theta) - y_{t,n}\|_2^2 \right\} \tag{4}$$

For the simplicity of explanation, we only consider one pair x and y in the following demonstration.

2.2. Forward-pass training subproblems

2.2.1. Feature generation

Unlike sparse representation, where each extracted image patch is approximated by a set of pre-trained bases, we use the equivalent convolution operation [17] and transfer the optimization of the bases into the network learning process. Therefore, the first layer of network can be described as follows

$$C_1 = \sigma(W_1 * x + b_1) \tag{5}$$

where W_1 denotes the convolution operator of size $c \times M_1 \times M_1 \times n_1$ and b_1 is the n_1 dimensional bias with its element associated with a filter. Here, c is the number of the image channels, M_1 means the filtered size and n_1 is the number of filters. We adopt the rectified linear unit (ReLu, $max(0, x)$) here for the nonlinear responses, which can be computed very efficiently [17].

2.2.2. Nonlinear mapping

We further perform non-linear mapping to project the n_{l-1} dimensional vectors into an n_l one, which is conceptually the refined feature and structure to represent the full-data-reconstructed image

$$C_l = \sigma(W_l * C_{l-1} + b_l) \tag{6}$$

where W_l is of a size $n_{l-1} \times M_l \times M_l \times n_l$.

2.2.3. *Last Layer convolution*

To produce the final predicted image from CNN, we explore another layer of convolution and hope to learn a set of linear filters W_L which are capable of projecting the coefficients onto the image domain

$$C_L = \sigma(W_L * C_{L-1} + b_L) \tag{7}$$

where W_L is of a size $n_{L-1} \times M_L \times M_L \times c$. To sum up, we have designed an L-layer convolutional neural network to learn the mapping relationship:

$$\begin{cases} C_0 = x \\ C_l = \sigma(W_l * C_{l-1} + b_l), l \in 1, 2, ..., L-1 \\ C_L = \sigma(W_L * C_{L-1} + b_L) \end{cases} \tag{8}$$

2.3. Backward Propagation

Given the training pair (x, y), the forward pass Eqs. (5-8) computes the activations and output values. To update the network parameters, as [16] we perform backward propagation to calculate the related gradients. Staring with the single pair objective, Eq. (4) can be written as

$$J(\Theta) = \underset{\Theta}{\operatorname{argmin}} \left\{ \frac{1}{2} \|C(x; \Theta) - y\|_2^2 \right\} \tag{9}$$

Let $D_l = W_l * C_{l-1} + b_l$ and δ^l denote the "error term" propagated backwards. We first calculate the last layer gradient

$$\delta^L = \frac{\partial J}{\partial b_L} = \frac{\partial J}{\partial D_L} \frac{\partial D_L}{\partial b_L} = C_L - y \tag{10}$$

Since $\frac{\partial D_l}{\partial b_l} = 1$ and $C_l = \sigma(D_l)$, δ^l of nonlinear mapping layer can be updated as follows

$$\delta^l = \frac{\partial J}{\partial b_l} = \frac{\partial J}{\partial D_{l+1}} \frac{\partial D_{l+1}}{\partial C_l} \frac{\partial C_l}{\partial D_l} = (\delta^{l+1} \star W^{l+1}) \circ \frac{\partial (D^l)}{\partial D^l} \tag{11}$$

where $\star$ means the cross-correlation operation which is different from the convolution in the feed-forward pass and $\circ$ denotes element-wise multiplication. Therefore, we can obtain the gradients for each layer

$$\begin{cases} \frac{\partial J}{\partial W_l} = \frac{\partial J}{\partial D^l} \frac{\partial D^l}{\partial W^l} = \delta^l \star D^{l-1} \\ \frac{\partial J}{\partial b_l} = \frac{\partial J}{\partial D^l} \frac{\partial D^l}{\partial b^l} = \frac{\partial J}{\partial D^l} = \delta^l \end{cases} \tag{12}$$

which can be used to calculate the stochastic gradients $\frac{\partial J(\Theta)}{\partial \Theta}$ during the training stage.

2.4. MR reconstruction formulation

Once we learned the hidden parameters $\hat{\Theta}$ from the pre-acquired datasets, we can reconstruct MR images by considering the following constrained optimization problem

$$\underset{u}{\operatorname{argmin}} \left\{ \|C(F^H f; \hat{\Theta}) - u\|_2^2 + \lambda \|f - F_M u\|_2^2 \right\} \tag{13}$$

As can be seen, this is a simple least squares problem admitting an analytical solution. And the least square solution satisfies the normal equation

$$(1 + \lambda F_M^H F_M) u = C(F^H f; \hat{\Theta}) + \lambda F_M^H f \tag{14}$$

By transforming the equation from image space to Fourier space, we have

$$(1 + \lambda F F_M^H F_M F^H) F u = F C(F^H f; \hat{\Theta}) + \lambda F F_M^H f \tag{15}$$

where $F F_M^H F_M F^H$ is a diagonal matrix consisting of ones and zeros. The ones are diagonal entries that correspond to the sampled locations in k-space. $F F_M^H f$ means zero-filled Fourier measurements. Therefore, we have

$$Fu(k_x, k_y) = \begin{cases} S(k_x, k_y) & , \text{ if } (k_x, k_y) \notin \Omega \\ \frac{S(k_x, k_y) + \lambda S_0(k_x, k_y)}{1 + \lambda} & , \text{ if } (k_x, k_y) \in \Omega \end{cases} \tag{16}$$

where Ω is the sampled location set.

2.5. Combination with CS-MRI reconstruction methods.

Besides the simple reconstruction model, we also provide two options for the integration with CS-MRI methods. a)Sequential model: Two-phase CS-MRI reconstruction. At the first stage, generate $C(F^H f; \hat{\Theta})$ from the learned network. At the second stage, initialize CS-MRI with $C(F^H f; \hat{\Theta})$ and then reconstruct MR images with CS-MRI. b) Integration model: Use the image generated by the network as a reference image and use it as additional regularization term.

$$\underset{u}{\operatorname{argmin}} \left\{ \|C(F^H f; \hat{\Theta}) - u\|_2^2 + \lambda \|f - F_M u\|_2^2 + \beta Reg(u) \right\} \tag{17}$$

where $Reg(u)$ is the sparse regularization term.

3. EXPERIMENTS AND RESULTS

Datasets: The training data consists of over 500 fully sampled MR brain images we collected from a 3T scanner (SIEMENS MAGNETOM TrioTim). The images are of a great diversity including axial, sagittal, horizontal ones, different contrast ones such as T1, T2 and PD-weighted images and of a variety of sizes. Informed consent was obtained from the imaging subject in compliance with the Institutional Review Board policy. Undersampled measurements were retrospectively obtained using the 1D low-frequency sampling mask and the 2D Poisson disk sampling mask. The large amount of corrupted/ground truth subimage pairs are then generated with the size of 33×33. Finally we use 90% of the subimage pairs as the training dataset and the rest 10% for validating the training process.

Implementation details: We use three layers of convolution for the network. The parameters are respectively set as $n_1 = 64, n_2 = 32, M_1 = 9, M_2 = 5$ and $M_3 = 5$. The filter weights of each layers are initialized by random values from a Gaussian distribution with zero mean and standard deviation 0.001. The bias are all initialized as 0. The training takes about three days, on a workstation equipped with 128G memory and a processor of 16 cores (Intel Xeon (R) CPU E5-2680 V3 @2.5GHz).

Figure 2 shows a set of reconstruction results of a transversal brain image. The brain dataset was obtained fully-sampled with 12-channel head coil and T2-weighted turbo spin-echo (TSE) sequence ($TE = 91.0ms, TR = 5000ms, FOV = 20 \times 20cm$, matrix = 256 $\times$270, slice thickness = $3mm$) via 3T scanner. And the data was then undersampled retrospectively with 1D low-frequency sampling mask at an acceleration factor of 3 and the 2D Poisson disk at an acceleration factor of 5. We also tested the proposed method on a sagittal brain image which was acquired on a GE 3T scanner (GE Healthcare, Waukesha, WI) with a 32-channel head coil and 3D T1-weighted spoiled gradient echo sequence (TE=minimum full, TR= 7.5ms, FOV=24 $\times$ 24 cm, matrix $= 256 \times 256$, slice thickness=1.7mm). We can observe from the images that there are quite a few details and structures captured by the network. Furthermore, the image generated by the simple reconstruction model is quite close to the original image. According to Fig. 3f, we can see the difference image is noise-like and consists only the contour information. There are no obvious details and structures lost. It demonstrates that the proposed network is capable of restoring the details and fine structures which are discarded in the zero-filled MR image. Furthermore, although the off-line training takes roughly three days, under the same GPU configurations, it takes far less than 1 second for each online MR reconstruction case.

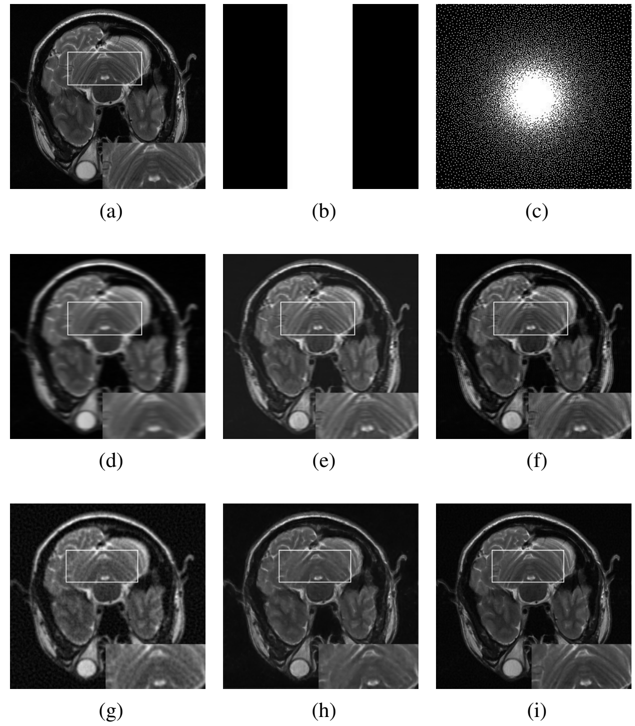

Fig. 2. (a) Ground truth image; (b) 1 D central low-frequency sampling mask with acceleration factor of 3; (c) 2D poisson undersampling mask with acceleration factor of 5; (d)(e)(f) the zero-filled MR image, network output and reconstruction result from 1D undersampled data; (g)(h)(i) the zero-filled M-R image, network output and reconstruction result from 2D undersampled data.

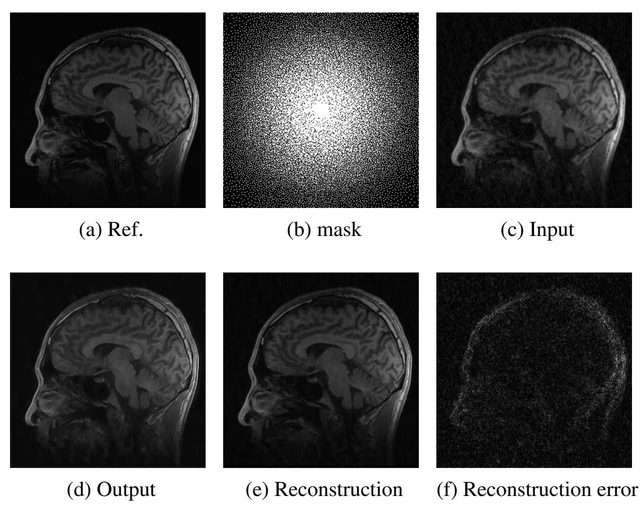

Fig. 3. The test results on another sagittal brain image at an acceleration factor of 3

4. CONCLUSIONS

An off-line convolutional neural network for accelerating M-R imaging is proposed in this paper, which includes the brief review and discussion of the concept, theoretical foundation, implementation and application of this network for undersampled MR image reconstruction. The experimental results on in-vivo MR images have shown the proposed network is capable of restoring the details and fine structures that are lost in the zero-filled MR image. We also provide two options for combing the proposed network with online CS-MRI methods for more efficient and effective imaging. More extensive experimental results will be provided in the future journal paper.

5. ACKNOWLEDGEMENT

We would like to thank Prof. Xiaogang Wang with The Chinese University of Hong Kong for his advice and helps regarding the CNN network.

6. REFERENCES

[1] M. Lustig and J. M. Pauly, “SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space,” *Magn.*

Reson. Med., vol. 64, no. 2, pp. 457–471, 2010.

[2] F. Knoll, C. Clason, K. Bredies, M. Uecker, and R. Stollberger, “Parallel imaging with nonlinear reconstruction using variational penalties,” *Magn. Reson. Med.*, vol. 67, no. 1, pp. 34–41, 2012.

[3] Yunmei Chen, William W. Hager, Feng Huang, Dzung T. Phan, Xiaojing Ye, and Wotao Yin, “Fast algorithms for image reconstruction with application to partially parallel MR imaging,” *SIAM J. Imaging Sciences*, vol. 5, no. 1, pp. 90–118, 2012.

[4] Xiaojing Ye, Yunmei Chen, and Feng Huang, “Computational acceleration for MR image reconstruction in partially parallel imaging,” *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1055–1063, 2011.

[5] Xiaojing Ye, Yunmei Chen, Wei Lin, and Feng Huang, “Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories,” *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 575–585, 2011.

[6] Xiaoqun Zhang, Martin Burger, Xavier Bresson, and Stanley Osher, “Bregmanized nonlocal regularization for deconvolution and sparse reconstruction,” *SIAM J. Imaging Sciences*, vol. 3, no. 3, pp. 253–276, 2010.

[7] Sajan Goud Lingala, Yue Hu, Edward DiBella, and Mathews Jacob, “Accelerated dynamic mri exploiting sparsity and low-rank structure: kt slr,” *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1042–1054, 2011.

[8] Jiao Wu, Fang Liu, LC Jiao, Xiaodong Wang, and Biao Hou, “Multivariate compressive sensing for image reconstruction in the wavelet domain: using scale mixture models,” *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3483–3494, 2011.

[9] Christian Wachinger, Mehmet Yigitsoy, Erik-Jan Rijkhorst, and Nassir Navab, “Manifold learning for image-based breathing gating in ultrasound and mri,” *Medical image analysis*, vol. 16, no. 4, pp. 806–818, 2012.

[10] Zhi-Pei Liang, Bruno Madore, Gary H Glover, and Norbert J Pelc, “Fast algorithms for gs-model-based image reconstruction in data-sharing fourier imaging,” *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 1026–1030, 2003.

[11] Shanshan Wang, Yong Xia, Qiegen Liu, Pei Dong, David Dagan Feng, and Jianhua Luo, “Fenchel duality based dictionary learning for restoration of noisy images,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5214–5225, 2013.

[12] Shanshan Wang, Qiegen Liu, Yong Xia, Pei Dong, Jianhua Luo, Qiu Huang, and David Dagan Feng, “Dictionary learning based impulse noise removal via L1-L1 minimization,” *Signal Processing*, vol. 93, no. 9, pp. 2696–2708, 2013.

[13] Qiegen Liu, Shanshan Wang, and Jianhua Luo, “A novel predual dictionary learning algorithm,” *J. Vis. Commun. Image Represent.*, vol. 23, no. 1, pp. 182–193, 2012.

[14] Qiegen Liu, Shanshan Wang, Kun Yang, Jianhua Luo, Yuemin Zhu, and Dong Liang, “Highly undersampled magnetic resonance image reconstruction using two-level Bregman method with dictionary updating,” *IEEE Trans. Med. Imaging*, vol. 32, no. 7, pp. 1290–1301, 2013.

[15] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, “3d convolutional neural networks for human action recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.

[16] Dan Ciresan, Ueli Meier, and Jürgen Schmidhuber, “Multi-column deep neural networks for image classification,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 3642–3649.

[17] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, “Learning a deep convolutional network for image super-resolution,” in *Computer Vision–ECCV 2014*, pp. 184–199. Springer, 2014.

[18] Junyuan Xie, Linli Xu, and Enhong Chen, “Image denoising and inpainting with deep neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 341–349.

[19] Justin Romberg, “Compressive sensing by random convolution,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.