ACCELERATING MAGNETIC RESONANCE IMAGING VIA DEEP LEARNING

Shanshan Wang¹, Zhenghang Su², Leslie Ying³, Xi Peng¹, Shun Zhu¹
Feng Liang⁴, Dagan Feng⁵ and Dong Liang¹

¹Paul C. Lauterbur Research Center for Biomedical Imaging, SIAT, CAS, Shenzhen, P.R.China
²School of Information Technologies, Guangdong University of Technology, Guangzhou, P.R. China
³Department of Biomedical Engineering and Department of Electrical Engineering, The State University of New York, Buffalo, New York 14260, USA
⁴Department of Industrial Engineering, NanKai University, Tianjin, P.R. China
⁵School 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 (CS) 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-
data like is generated as the direct inverse transform of the observed Hermitian transpose operation. The zero-filled MR image \( z \) as \( F^H F \) mask, where \( P \) denotes the full Fourier encoding matrix normalized \( F^H F = I \), \( u \) is the original (ground truth) image and therefore \( F u \) 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

\[
z = F^H P F u
\]

As stated in [19], in terms of linear algebra, the circular convolution of a signal \( u \) with a pulse \( p \) can be written as \( F^H P F u \), 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

\[
\arg \min_{\Theta} \left\{ \frac{1}{2T} \sum_{t=1}^{T} \| C(z_t; \Theta) - u_t \|^2 \right\}
\]

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

\[
C_1 = \sigma(W_1 * x + b_1)
\]

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

\[
C_l = \sigma(W_l * C_{l-1} + b_l)
\]

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

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

(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:

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

(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

$$J(\Theta) = \arg\min_{\Theta} \left\{ \frac{1}{2} ||C(x; \Theta) - y||_2^2 \right\}$$

(9)

Let $D_l = W_l \ast C_{l-1} + b_l$ and $\delta^l$ denote the “error term” propagated backwards. We first calculate the last layer gradient

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

(10)

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

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

(11)

where $\ast$ 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

$$\begin{cases}
\frac{\partial J}{\partial W^l} = \frac{\partial J}{\partial D^l} \frac{\partial D^l}{\partial W^l} = \delta^l \ast D^{l-1} \\
\frac{\partial J}{\partial b^l} = \frac{\partial J}{\partial D^l} \frac{\partial D^l}{\partial b^l} = \frac{\partial J}{\partial D^l} = \delta^l
\end{cases}$$

(12)

which can be used to calculate the stochastic gradients $\frac{\partial J}{\partial \Theta}$ 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

$$\arg\min_u \left\{ ||C(F^H f; \hat{\Theta}) - u||_2^2 + \lambda ||f - F_M u||_2^2 \right\}$$

(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

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

(14)

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

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

(15)

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

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

(16)

where $\Omega$ 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.

$$\arg\min_u \left\{ ||C(F^H f; \hat{\Theta}) - u||_2^2 + \lambda ||f - F_M u||_2^2 + \beta \text{Reg}(u) \right\}$$

(17)

where $\text{Reg}(u)$ 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 $33 \times 33$. 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 = 91.0 ms, TR = 5000 ms, FOV = 20 × 20 cm, matrix = 256 × 270, slice thickness = 3 mm) 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=7.5 ms, FOV=24 × 24 cm, matrix = 256 × 256, slice thickness=1.7 mm). 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 offline training takes roughly three days, under the same GPU configurations, it takes far less than 1 second for each online MR reconstruction case.

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


