



# Generalized Deep Learning-Based Proximal Gradient Descent for MR Reconstruction

Guanxiong Luo<sup>1,2</sup>, Mengmeng Kuang<sup>1</sup>, and Peng Cao<sup>1</sup>(✉)

<sup>1</sup> The University of Hong Kong, HKSAR, Pok Fu Lam, China  
caopeng1@hku.hk

<sup>2</sup> University Medical Center Göttingen, Göttingen, Germany

**Abstract.** The data consistency for the physical forward model is crucial in inverse problems, especially in MR imaging reconstruction. The standard way is to unroll an iterative algorithm into a neural network with a forward model embedded. The forward model always changes in clinical practice, so the learning component's entanglement with the forward model makes the reconstruction hard to generalize. The deep learning-based proximal gradient descent was proposed and use a network as regularization term that is independent of the forward model, which makes it more generalizable for different MR acquisition settings. This one-time pre-trained regularization is applied to different MR acquisition settings and was compared to conventional  $\ell_1$  regularization showing  $\sim 3$  dB improvement in the peak signal-to-noise ratio. We also demonstrated the flexibility of the proposed method in choosing different undersampling patterns.

**Keywords:** Magnetic Resonance Imaging · Image reconstruction · Deep Learning · Proximal gradient descent · Learned regularization term

## 1 Introduction

Before employing deep learning in accelerated MRI reconstruction, conventional methods for parallel MR imaging are based on the numerical pseudo-inversion of ill-posed MRI encoding matrix, which could be prone to reconstruction error at poor conditioning [4, 6, 7]. The encoding matrix comprises the k-space under-sampling scheme, coil sensitivities, Fourier transform. The traditional reconstruction involves some gradient descent methods for minimizing the cost function of the k-space fidelity and the regularization term [6, 9]. There is a tradeoff between the image artifact level and the under-sampling rate as limited by the encoding capacity of coil sensitivities. Nevertheless, the parallel imaging technique robustly provides the acceleration at factor 2–4 [4, 7]. The compressed sensing technique exploits the sparsity property of MR images in a specific transform domain, such as the wavelet domain, in combination with the incoherent under-sampling in k-space, which enables even larger acceleration factor.

With the fast growth of machine learning, the supervised learning have been applied to MRI reconstruction [1, 8, 11]. Those methods MRI encoding matrices were fully

---

P. Cao—Department of Diagnostic Radiology, LKS Faculty of Medicine, Pokfulam Road, HK, SAR, China.

included in the neural network models. These models were trained with predetermined encoding matrices and corresponding under-sampling artifacts. After training, imaging configurations, including the k-space under-sampling schemes and coil sensitivities, associated encoding matrices, must also be unchanged or changed only within predetermined sampling patterns, during the validation and application, which could be cumbersome or to some extent impractical for the potential clinical use.

To tackle this design challenge, we unroll proximal gradient descent steps into a network and call it Proximator-Net. Inspired by [3], the proposed method was adapted from proximal gradient descent. This study’s objective was to develop a flexible and practical deep learning-based MRI reconstruction method and implement and validate the proposed method in an experimental setting regarding changeable k-space under-sampling schemes.

## 2 Method - Proximal Network as a Regularization Term

The image reconstruction for MR k-space data acquired with different k-space trajectories can be formulated as an inverse problem for the corresponding forward model. We followed the MRI reconstruction problem formulation used in  $l_1$ -ESPIRiT [9]. Let a forward operator  $\mathcal{A}$  to map the MR image  $\mathbf{x}$  to the sampled k-space data  $\mathbf{y}$ . The operator  $\mathcal{A}$  consists of a Fourier transform operator  $\mathcal{F}$ , coil sensitivity  $\mathcal{S}$ , and a k-space sampling mask  $\mathcal{P}$ , i.e.,  $\mathcal{A} = \mathcal{P}\mathcal{F}\mathcal{S}$  [9]. The well-known solution to the inverse problem can be formulated as an optimization problem with a regularization term [9] as

$$\bar{\mathbf{x}} = \arg \min_x \|\mathcal{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}), \quad (1)$$

where  $\phi(\mathbf{x})$  is a regularization term that introduces the prior knowledge, and  $\lambda$  the regularization parameter. The proximal operator for  $\phi(\mathbf{x})$  in the proposed approach,  $\mathbf{prox}_\phi(\mathbf{x})$ , is defined as:

$$\hat{\mathbf{x}} = \mathbf{prox}_{\phi, 1/\lambda}(\mathbf{v}) := \arg \min_{\mathbf{x}} \|\mathbf{x} - \mathbf{v}\|_2^2 + \lambda\phi(\mathbf{x}) \quad (2)$$

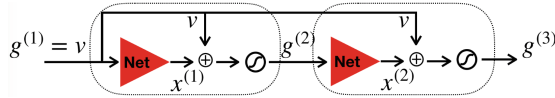
$$\hat{\mathbf{x}} \in \chi \sim P_{model} \approx P_{data} \quad (3)$$

where  $\hat{\mathbf{x}}$  is the proximate value at  $\mathbf{v}$ ,  $\chi$  is the sub-space that contains MR images (including the ground truth),  $P_{data}$  is the probability distribution of the observed MR images, and  $P_{model}$  is learned the probability distribution. Taking the sparse constraint in compressed sensing as an analogy,  $\phi(\mathbf{x})$  is a  $l_1$ -norm of coefficients from the wavelet transform, the Equation (2) can be approximated by the shrinkage method (FISTA). Back to the definition of the proximal operator, we treat Equation (2) as another optimization problem solved iteratively by the proximal gradient descent method, which is expressed as follows

$$\mathbf{g}^{(t+1)} = \mathbf{x}^{(t)} + \alpha^{(t)}(\mathbf{v} - \mathbf{x}^{(t)}) \quad (4a)$$

$$\mathbf{x}^{(t+1)} = \mathbf{Net}_\phi(\mathbf{g}^{(t+1)}) \quad (4b)$$

then, we unroll above iterative steps as a Proximator-Net shown in Fig. 1.



**Fig. 1.** The unrolled proximate gradient descent as neural network, the  $\text{Net}_\phi$  are the red triangles whose detailed structure is detailed in material.

At the iteration  $t$ ,  $\mathbf{x}^{(t+1)}$  is supposed be closer to  $P_{data}$  than  $\mathbf{x}^{(t)}$ . In order to achieve this, images corrupted by Gaussian noises are used as input. The corresponding noise-free images are used as labels. The  $\ell_2$ -loss between outputs and labels is used. To solve Equation (3) and make the prediction close to the data distribution, the  $\ell_2$  regularization of the gradient of output with respect to input is used, as proposed in [2]. Therefore the final training loss is

$$\mathbb{E} \left[ \left\| r_\theta(\tilde{\mathbf{x}}) - \mathbf{x} \right\|_2^2 + \sigma^2 \left\| \frac{\partial r_\theta(\mathbf{x})}{\partial \mathbf{x}} \right\|_2^2 \right], \tag{5}$$

where  $\tilde{\mathbf{x}}$  is the image perturbed by the Gaussian noise with  $\sigma$  and  $r_\theta(\mathbf{x})$  is the neural network parameterized by  $\theta$ . Proximator-Net is recast as a neural network-based proximal operator for the proposed regularization term, and the proposed approach for reconstruction is

$$\mathbf{m}^{(t)} = \mathbf{x}^{(t)} + \alpha^{(t)} \mathcal{A}^H(\mathbf{y} - \mathcal{A}\mathbf{x}^{(t)}) \tag{6a}$$

$$\mathbf{x}^{(t+1)} = (1 - \lambda) \cdot \mathbf{x}^{(t)} + \lambda \cdot r_\theta(\mathbf{m}^{(t)}) \tag{6b}$$

where  $\lambda$  is the parameter for the learned regularization term.

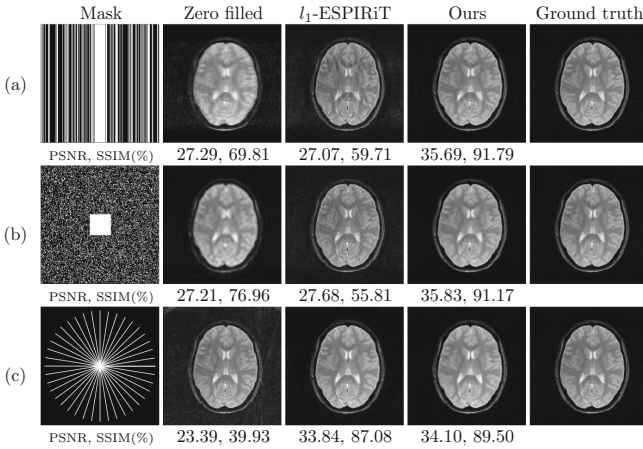
### 3 Experiments and Results

**Dataset and Pre-processing:** The dataset we used in this work is from Ref. [5]. Training images were reconstructed from 8 channels k-space data without undersampling. Then, these image data-sets after coil combination were scaled to a magnitude range of  $[-1, 1]$  and resized to an image size of  $256 \times 256$ . In the end, 900 images were used for training, and 300 images were used for testing. We set the level of Gaussian noise with  $\sigma = 0.03, \mu = 0$ . The training was performed with Tensorflow on 4 GTX 2080Ti GPUs and took about 2 h for 500 epochs. Real and imaginary parts of all 2D images were separated into two channels when inputted into the neural network.

**Table 1.** Comparison of PSNRs (in dB, mean  $\pm$  standard deviation,  $N = 300$ ) and SSIMs(%) between  $\ell_1$ -ESPIRiT and the proposed method.

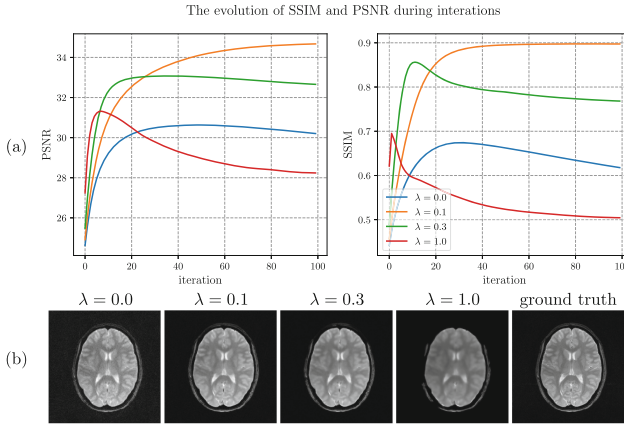
Sampling rate	$\ell_1$ -ESPIRiT		Ours	
	PSNR(dB)	SSIM(%)	PSNR(dB)	SSIM(%)
<i>1D mask</i>				
20%	31.63 $\pm$ 5.34	68.28 $\pm$ 14.48	35.12 $\pm$ 3.41	91.34 $\pm$ 5.32
30%	34.25 $\pm$ 5.66	76.31 $\pm$ 12.54	37.96 $\pm$ 3.58	93.87 $\pm$ 4.37
<i>2D mask</i>				
20%	34.73 $\pm$ 5.84	71.48 $\pm$ 16.06	37.90 $\pm$ 2.86	93.55 $\pm$ 4.09
30%	37.55 $\pm$ 5.34	80.95 $\pm$ 11.84	40.12 $\pm$ 2.98	95.09 $\pm$ 3.41

**Reconstruction and Comparisons:** The proposed method in Eq. 6 was implemented with Python. We performed  $\ell_1$ -ESPIRiT (regularization parameter=0.005) reconstruction with the BART toolbox [9, 10]. The coil sensitivity map was estimated from a  $20 \times 20$  calibration region in the central k-space using ESPIRiT. To compare the images reconstructed using different methods, we calculated the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) with respect to ground truth. Also, we validated the proposed method in different acquisition settings such as acceleration along one dimension, two dimensions, or radial acquisition. The metrics of different reconstructions were shown in Table 1. Generally, the result demonstrated that the proposed method could restore the high-quality MRI with both high acceleration factor (i.e., 20% samples) and high PSNR (i.e.,  $>35$  dB). Figure 2 shows the comparison between  $\ell_1$ -ESPIRiT and the proposed method.



**Fig. 2.** Row (a) shows the proposed method can remove the aliasing artifacts along phase encoding direction in the case of using 30% k-space. Row (b) shows the proposed method can eliminate the burring fog on the image in the case of using 20% k-space. Row (c) shows the two methods' performances were close. With 40 radial spokes acquired in k-space.

**Smoothing Effect of the Learned Regularization Term:** The smoothing effect of the proposed regularization term was shown in Fig. 3. The parameter  $\lambda$  was tunable in Equation (6b). When  $\lambda$  was set to be zero, it was an iterative parallel imaging (i.e., SENSE) reconstruction [7]. With the increase of  $\lambda$ , the artifacts and noises disappeared, and also the image appeared more smooth with some details eclipsed. We noticed that setting  $\lambda$  around 0.1 provided better PSNR and SSIM. We plotted SSIM and PSNR over iterations to monitor the quality of reconstruction, using varied tuning parameter  $\lambda$ .



**Fig. 3.** (a) The curve of SSIM and PSNR over iterative steps. (b) Images reconstruction from 40 radial k-space spokes with different  $\lambda$ .

## 4 Discussion and Conclusion

In this study, a once pre-trained neural network - Proximator-Net - is used as a regularization in iterative algorithm for MRI reconstruction and can be applied to different reconstruction tasks, taking advantage of the separation of the learned information from the fidelity enforcement. That makes it different to previous methods [1, 8, 11]. In this initial experiment, we focused on demonstrating the utility of the proposed method in classic compressed sensing and radial k-space acquisition, and we used the brain MRI data to evaluate the method. Like conventional iterative reconstruction algorithms, k-space fidelity in the proposed hybrid approach was enforced by the least-square term and implemented outside the neural network, allowing the high flexibility to change k-space under-sampling schemes and RF coil settings. For quantitative comparison, our methods achieved 3dB higher PSNR in the tested acquisition settings compared with  $l_1$ -ESPIRiT.

## References

1. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imaging* **38**(2), 394–405 (2018)
2. Alain, G., Bengio, Y.: What regularized auto-encoders learn from the data-generating distribution. *J. Mach. Learn. Res.* **15**(1), 3563–3593 (2014)
3. Chen, D., Davies, M.E., Golbabaee, M.: Compressive mr fingerprinting reconstruction with neural proximal gradient iterations (2020)
4. Griswold, M.A., et al.: Generalized autocalibrating partially parallel acquisitions (grappa). *Magn. Reson. Med.* **47**(6), 1202–1210 (2002)
5. Luo, G., Zhao, N., Jiang, W., Hui, E.S., Cao, P.: MRI reconstruction using deep bayesian estimation. *Magn. Reson. Med.* **84**(4), 2246–2261 (2020)
6. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: the application of compressed sensing for rapid mr imaging. *Magn. Reson. Med.* **58**(6), 1182–1195 (2007)
7. Pruessmann, K.P., Weiger, M., Börnert, P., Boesiger, P.: Advances in sensitivity encoding with arbitrary k-space trajectories. *Magn. Reson. Med.* **46**(4), 638–651 (2001)
8. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE T.M.I.* **37**(2), 491–503 (2017)
9. Uecker, M., Lai, P., Murphy, M.J., Virtue, P., Elad, M., Pauly, J.M., Vasanawala, S.S., Lustig, M.: *Espirit* an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. *Magn. Reson. Med.* **71**(3), 990–1001 (2014)
10. Uecker, M., et al.: Berkeley advanced reconstruction toolbox. In: *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 23, p. 2486 (2015)
11. Yang, Y., Sun, J., Li, H., Xu, Z.: Deep admm-net for compressive sensing mri. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) *Advances in Neural Information Processing Systems* 29, pp. 10–18. Curran Associates, Inc. (2016)