

# Image Reconstruction by Splitting Deep Learning Regularization from Iterative Inversion

Jiulong Liu, Tao Kuang, and Xiaoqun Zhang

Institute of Natural Sciences & School of Mathematical Sciences & MOE-LSC  
Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China 200240  
xqzhang@sjtu.edu.cn

**Abstract.** Image reconstruction from downsampled and corrupted measurements, such as fast MRI and low dose CT, is mathematically ill-posed inverse problem. In this work, we propose a general and easy-to-use reconstruction method based on deep learning techniques. In order to address the intractable inversion of general inverse problems, we propose to train a network to refine intermediate images from classical reconstruction procedure to the ground truth, i.e. the intermediate images that satisfy the data consistence will be fed into some chosen denoising networks or generative networks for denoising and removing artifact in each iterative stage. The proposed approach involves only techniques of conventional image reconstruction and usual image representation/denoising deep network learning, without a specifically designed and complicated network structures for a certain physical forward operator. Extensive experiments on MRI reconstruction applied with both stack auto-encoder networks and generative adversarial nets demonstrate the efficiency and accuracy of the proposed method compared with other image reconstruction algorithms.

**Keywords:** Inverse problems; MRI reconstruction; Deep Learning; Regularization

## 1 Introduction

Image reconstruction problems arisen in medical imaging area such as fast MRI and low dose CT are mathematically ill-posed inverse problems. We often consider a linear imaging system with a forward operator  $A$ , for example partial 2D Fourier transform for MRI and X-ray transform for CT. The measurement  $y$  is given as  $y = Ax$  for  $x$  being the underlying image in the perfect noise free case. The linear operator  $A$  is ill-posed for most applications; therefore some statistical priors are necessary to make these problems invertible. Sparsity priors such as total variation (TV) [1] and wavelet tight frame [2] have been among those popular regularization and studied extensively in the literature.

In practice, the measurements are often corrupted by noise, i.e.

$$y = Ax + \epsilon \tag{1}$$

if we assume it is i.i.d additive Gaussian noise. Derived by a maximum-likelihood estimator of the physical process and sparsity prior distribution of the original image, it is common to solve the following unconstrained model

$$\min_x \frac{1}{2} \|Ax - y\|_2^2 + \mu \|Dx\|_1, \quad (2)$$

where  $\mathcal{D}$  is a sparsity transform, for example, the mentioned spatial gradient operator  $\nabla$  (total variation) and a tight frame transform  $\mathcal{W}$ .

However, some side effects will also be involved by sparsity regularization due to the predefined sparsity transform, for example, the staircasing effects are introduced by TV. Deep networks have been successfully applied to many image restoration tasks such as image denoising, inpainting, super resolution [3–5]. It is shown in these work that those delicately designed deep networks achieved state-of-the-art performance for these image processing problems. However, despite their superior performance, it is still challenging to adapt the network for medical image reconstruction problems as the networks are specifically designed for those particular forward operators. Most of the emerged deep learning based medical image reconstruction are based on the sparsity optimization algorithms such as primal dual methods and Alternating Direction of Multiplier methods (ADMM). For example, ADMM-net [6] and the learned variational network in [7] aim to mimic the optimization algorithms for solving the sparse regularization model (2) and build a network to learn the sparsity transform  $\mathcal{D}$ . In [8, 6, 9], analytic solutions are obtained for the inversion layers and a proximal operator is learned for the denoising/anti-artifact layers. In the work [10, 11], the authors carefully designed a MRI reconstruction network to enhance data consistency. These networks achieve state-of-the-art reconstruction results and at the same time are usually more complicated compared to common neural networks, especially for derivative computing.

Because of the intractability of inversion of an ill-posed operator with partial and corrupted measurements, we do not intend to learn an end-to-end inversion mapping from the measurements to the reconstructed image as previous work. Inspired by regularization based image reconstruction methods, we propose to split the task of inversion of a known forward operator from learning an image representation network. In order to feed the inputs into networks implicitly, we establish a data consistency constrained network loss function and then apply ADMM to split the tasks of solving the inversion and learning a network. The problem is solved through simple iterations of existing techniques of conventional inversion and usual image representation/denoising deep network learning. We note that our method is different from ADMM-net, as ADMM-net considered the solution of the sparsity optimization algorithms ADMM as the network output and the sparsity transform  $\mathcal{D}$  is considered as network parameters to be learned. Our method does not intend to design a new network structure but integrate existing ones in the ADMM algorithm to solve the proposed model. The prior of to-be-reconstructed images is obtained by the learned network, which can be easily used for the inference process. Finally, data consistency is maintained

through iteration for the reconstruction, which is usually not the case for most of learning based reconstruction methods.

## 2 Our approach

The Stacked Auto-Encoder (SAE) deep network has shown to be a useful image representation method for denoising [12], where a greedy layerwise approach is proposed for pretraining. Stacked Convolutional Auto-Encoder (SCAE) [13] was further proposed to preserve frequently repeated local image features. And some improvement has been achieved by the Stacked Non-Local Auto-Encoder (SNLAE) in [14] by using a nonlocal collaborative stabilization. In recent years, more and more networks emerge for image restoration problems. For example, it has been demonstrated that generative adversarial network (GAN) model is powerful for medical or natural image restoration problems such as super-resolution [4] and deburring [15].

In the following, we propose our image reconstruction learning model based on a denoising network or GAN model. We denote the input dataset for a network  $\mathbf{x} = \{\mathbf{x}_k\}_{k=1}^m$  with the corresponding ground truth  $\tilde{\mathbf{x}} = \{\tilde{\mathbf{x}}_k\}_{k=1}^m$  where  $m$  is the number of samples. In image reconstruction inverse problems, we denote the corresponding measurements  $\mathbf{y} = \{\mathbf{y}_k\}_{k=1}^m$  for  $\mathbf{y}_k = A\mathbf{x}_k$  where  $A$  is a known forward operator. Here we use the boldface to denote the vectors of all the input and output images in the training procedure and we use the regular characters for their counterparts for the inference.

The learning procedure of a denoising network is designed to minimize a cost function  $L_H(\mathbf{x}, \theta)$ , for example the quadratic function

$$\min_{\theta} L_H(\mathbf{x}, \theta) := \|f(\mathbf{x}, \theta) - \tilde{\mathbf{x}}\|_2^2 \quad (3)$$

where  $f(\mathbf{x}, \theta)$  is the output and  $\theta$  is the set of network parameters. For GAN model, the following min-max problem is considered

$$\max_{\theta_d} \min_{\theta_g} L_G(\mathbf{x}, \theta_g, \theta_d) = \frac{1}{m} \sum_{i=1}^m [\log(D(\tilde{\mathbf{x}}_i, \theta_d))] + \frac{1}{m} \sum_{i=1}^m [\log(1 - D(G(\mathbf{x}_i, \theta_g), \theta_d))] \quad (4)$$

where  $\theta_g$  and  $\theta_d$  are the parameter sets for the generative and discriminative networks respectively, and  $G(\cdot, \theta_g)$  and  $D(\cdot, \theta_d)$  are the outputs of the two networks.

Let

$$J(\mathbf{x}) = \eta \left( \sum_{i=1}^m \|A\mathbf{x}_i - \mathbf{y}_i\|_2^2 + \mu \|D\mathbf{x}_i\|_1 \right) \quad (5)$$

be the conventional data consistency term with sparse regularization. The formulation can be easily generalized for other data fidelity derived from maximum likelihood of a posteriori estimation, and with other regularization term in  $J(\mathbf{x})$ . The regularization parameter  $\mu$  can be very small or even zero.

Being motivated by the fact that many powerful networks are available for removing noise and artifact, we now attempt to propose to integrate deep learning network in the image reconstruction. Our basic idea is to use the variable  $\mathbf{x}$  which meets the data consistence implicitly to fed into the to-be-learned networks, by solving the following problem with a deep learning regularization

$$\min_{\mathbf{x}, \theta} L_H(\mathbf{x}, \theta) + J(\mathbf{x}), \quad (6)$$

and

$$\max_{\theta_d} \min_{\theta_g, \mathbf{x}} L_G(\mathbf{x}, \theta_g, \theta_d) + J(\mathbf{x}) \quad (7)$$

for  $L_H$  and  $L_G$  being the cost function for a denoising network and GAN model respectively.

The above two models can be solved by adapting ADMM algorithm [16]. Taking (6) as an example, we reformulate it as

$$\begin{aligned} \min_{\mathbf{x}, \theta, \mathbf{z}} \quad & L_H(\mathbf{x}, \theta) + J(\mathbf{z}) \\ \text{s.t.} \quad & \mathbf{x} = \mathbf{z}. \end{aligned} \quad (8)$$

The augmented Lagrangian for the problem (8) is given as

$$L_\rho(\mathbf{x}, \theta, \mathbf{z}, \mathbf{p}) = L_H(\mathbf{x}, \theta) + J(\mathbf{z}) + \mathbf{p}^T(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \quad (9)$$

for a parameter  $\rho > 0$ .

The idea of the ADMM algorithm for solving the optimization problem (8) is to alternately update the primal variables  $\mathbf{x}, \theta, \mathbf{z}$  by minimizing the augmented Lagrangian function (9) and update the dual variable  $\mathbf{p}$  with a dual ascent step, which leads to the following scheme

$$\begin{cases} \mathbf{z}^{k+1} = \arg \min_{\mathbf{z}} J(\mathbf{z}) + \frac{\rho}{2} \|\mathbf{x}^k - \mathbf{z} + \mathbf{b}^k\|_2^2 \\ \theta^{k+1} = \arg \min_{\theta} L_H(\mathbf{x}^k, \theta) \\ \mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} L_H(\mathbf{x}, \theta^{k+1}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}^{k+1} + \mathbf{b}^k\|_2^2 \\ \mathbf{b}^{k+1} = \mathbf{b}^k + (\mathbf{x}^{k+1} - \mathbf{z}^{k+1}) \end{cases} \quad (10)$$

for  $\mathbf{p}^k = \rho \mathbf{b}^k$ . The variables  $\mathbf{x}^0$  and  $\mathbf{z}^0$  are initialized by

$$\mathbf{x}^0 = \mathbf{z}^0 = \arg \min_{\mathbf{z}} J(\mathbf{z}) . \quad (11)$$

For the first subproblem in (10), we can solve this conventional reconstruction problem with a classical reconstruction method, such as ADMM again if there is a sparse regularization term present in  $J$ ; For the second subproblem in (10), it is a typical loss function minimization for a deep learning network with the input  $\mathbf{x}^k$ , and a stochastic gradient descent method built in the neural network tools can be applied; For the third subproblem in (10), we can also use a stochastic gradient descent method.

The similar alternating scheme as (10) can be obtained for solving the GAN training model (7) by replacing the second step in (10) by

$$(\theta_g^{k+1}, \theta_d^{k+1}) = \arg \max_{\theta_d} \min_{\theta_g} L_G(\mathbf{x}^k, \theta_g, \theta_d). \quad (12)$$

Here we need to alternately apply gradient descent for updating  $\theta_g^{k+1}$  and gradient ascent for updating  $\theta_d^{k+1}$  as the general GAN methods do.

After we obtain the network parameter set  $\theta^*$ , the learned network is ready to be used for mapping a given input image  $\hat{x}$  to an estimated ground truth image  $x$  by  $x = f(\hat{x}, \theta^*)$ . More precisely, given a measurement  $y$ , we obtain the reconstructed image  $x$  through the following scheme

$$\begin{cases} z^{k+1} = \arg \min J(z) + \frac{\rho}{2} \|x^k - z + b^k\|_2^2 \\ \hat{x}^{k+1} = f(x^k, \theta^*) \\ x^{k+1} = \frac{1}{1+\rho} (\hat{x}^{k+1} + \rho(z^{k+1} - b^k)) \\ b^{k+1} = b^k + (\hat{x}^{k+1} - z^{k+1}). \end{cases} \quad (13)$$

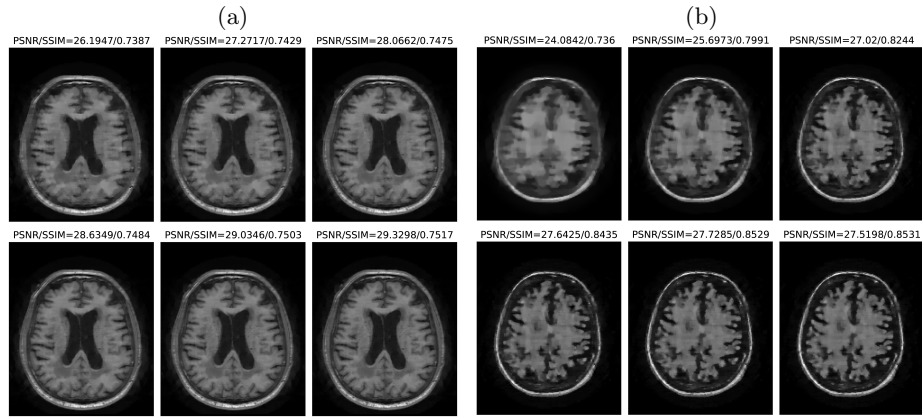
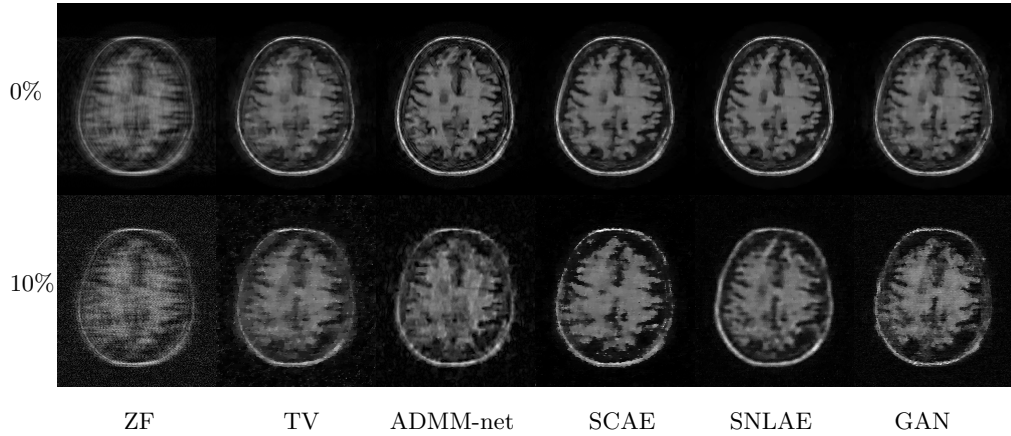
The initialization of  $\mathbf{x}^0$  and  $\mathbf{z}^0$  are performed similarly as (11). For the GAN based reconstruction model, we can use the similar scheme to (13) by replacing  $f(\mathbf{x}^{k+1}, \theta^*)$  with  $G(\mathbf{x}^{k+1}, \theta_g^*)$  in the second step to obtain the reconstructed image  $x$  from a measurement.

### 3 Experiments

In this section, we perform the experiments on MRI reconstruction from down-sampled measurements. The MRI data are generated by partial Fourier transform with Gaussian noise corruption, i.e.  $y = K\mathcal{F}(x + l * (\xi_1 + \xi_2 * i))$  where  $l$  is the noise level,  $\xi_1, \xi_2$  obey i.i.d normal distribution,  $x$  is the ground truth image, and  $K$  is the downsample operator. In our experiments, the MRI image dataset is from ADNI (Alzheimer’s Disease Neuroimaging Initiative) of which 300 slices of size  $192 \times 160$  are used for training and 21 slices are used for inferring, and three different downsampling patterns with three downsampling rates are used for simulating the measurements. To speed up the training process and alleviate the ill-conditionness when the sampling rate is severely low, we use TV term in the reconstruction functional  $J(x)$  (5), but its weight  $\mu$  is decreasing by outer loops of our method. As  $\mu$  gets smaller and smaller, the contribution of TV is eventually much smaller than what is used for sparsity regularized reconstruction. In order to demonstrate the flexibility of our approach, we implement three kinds of networks for MRI reconstruction, i.e. SCAE [13], SNLAE [14] and GAN [5]. The basic setup and training/test time on a PC with an Intel i7 and a Nvidia GPU GTX1060 for the three networks SCAE, SNLAE and GAN are listed in Tab. 1. To demonstrate the convergence of the proposed method, the intermediate training results and inferring results by SCAE of one slice are shown in Fig. 1, in which we can see that the images trend to be of good quality and higher PSNR.

**Table 1.** Network setup and computation time

Parameters	SCAE/GAN Gen.	SNLAE	GAN Discri.
Dataset size	(300, 192, 160)	(300, 192, 160)	(300, 192, 160)
Number of layers	6	6	8
Filter size in hidden layers	[3, 3, 3, 3, 3]	–	[3, 3, 3, 3, 3, 1, 1]
Stride in hidden layers	[1, 1, 1, 1, 1]	–	[2, 2, 2, 2, 2, 1, 1]
Number of neurons/filters in hidden layers	$[2^5, 2^6, 2^7, 2^6, 2^5]$	$[2^9, 2^{10}, 2^{11}, 2^{10}, 2^9]$	$[2^3, 2^6, 2^6, 2^6, 2^6, 2^6, 2^6]$
Outer loops	5	5	5
Train time(h)	11	16	36
Test time(sec)	3.2/3.3	4.1	–

**Fig. 1.** Intermediate results  $[f(x^0, \theta^0), f(x^1, \theta^1), \dots, f(x^5, \theta^5)]$  for MRI reconstruction with 25% radial downsampling by SCAE. (a) training step; (b) inferring step**Fig. 2.** MRI reconstruction results. Sampling pattern and rate: 1D random with 25%; The first row: noise free; The second row: 10% noise.

In Fig. 2, we show the comparison of our methods with the zero-filling method (ZF) [17], TV regularization based reconstruction [18], and ADMM-net [6] for the case with 1D random downsampling pattern and 25% downsampling rate (the results for the other sampling patterns and rates are provided in the Supplementary file). For the three downsampling patterns, we observe that in the case of noise-free, the reconstructed images by our proposed methods including SCAE, SNLAE and GAN, and by ADMM-net have better spatial resolution. In the case of noise level 10% and with measurements of very low sampling rate, our method with the four networks still achieve good performance. Specially, for the case with 1D random downsampling pattern, we can find our methods have alleviate both noise and artifact. Visually, our methods generally achieved cleaner images compared to ADMM-net for noisy data. To assess reconstruction image quality quantitatively, we further show the results of PSNR and SSIM in Tab. 2. All the best PSNR are scattered at the methods with learned regularization.

**Table 2.** MRI reconstruction quality(PSNR/SSIM)

Method	Noise	Rate	ZF	TV	ADMM-net	SCAE	SNLAE	GAN
radial	0%	1/5	24.36/0.47	30.73/0.86	<b>32.31</b> /0.92	32.00/0.92	30.47/0.83	30.13/0.84
		1/4	25.45/0.51	32.32/0.90	33.67/0.93	<b>33.94</b> /0.94	32.53/0.88	32.26/0.90
		1/3	27.25/0.56	34.60/0.94	35.27/0.94	<b>36.37</b> /0.96	35.15/0.92	34.49/0.94
	10%	1/5	22.18/0.35	24.69/0.49	25.44/0.59	25.52/0.73	<b>25.98</b> /0.68	25.02/0.73
		1/4	22.38/0.36	25.16/0.49	25.96/0.61	26.13/0.70	<b>26.38</b> /0.66	25.53/0.74
		1/3	22.37/0.37	25.28/0.49	26.50/0.60	26.64/0.74	26.70/0.65	<b>26.71</b> /0.75
2D random	0%	1/5	24.91/0.49	31.69/0.89	33.81/0.93	<b>34.24</b> /0.94	31.95/0.86	31.79/0.89
		1/4	25.30/0.50	32.79/0.90	34.97/0.94	<b>35.61</b> /0.95	32.85/0.86	32.94/0.91
		1/3	26.32/0.53	34.93/0.93	36.31/0.95	<b>37.71</b> /0.96	35.33/0.91	35.10/0.94
	10%	1/5	22.37/0.37	24.97/0.51	25.42/0.61	<b>25.90</b> /0.73	25.97/0.67	25.78/0.75
		1/4	22.38/0.36	24.92/0.49	25.84/0.60	26.06/0.74	26.15/0.67	<b>26.31</b> /0.75
		1/3	22.37/0.37	24.91/0.47	26.14/0.56	26.38/0.72	26.41/0.62	<b>26.48</b> /0.76
1D random	0%	1/5	22.78/0.61	25.22/0.75	28.53/0.85	<b>28.79</b> /0.87	28.73/0.86	27.21/0.81
		1/4	23.06/0.62	25.77/0.76	28.99/0.87	<b>29.37</b> /0.88	29.06/0.86	27.47/0.82
		1/3	23.86/0.65	27.34/0.81	<b>32.18</b> /0.91	31.25/0.91	30.98/0.89	30.09/0.86
	10%	1/5	20.72/0.27	22.38/0.39	22.59/0.40	22.22/0.61	<b>24.52</b> /0.60	22.76/0.67
		1/4	20.37/0.26	22.25/0.37	22.98/0.44	22.72/0.63	<b>24.39</b> /0.56	23.32/0.69
		1/3	20.37/0.28	22.59/0.37	23.96/0.47	23.75/0.62	<b>24.98</b> /0.58	23.93/0.70

## 4 Conclusion

We developed a variational image reconstruction method which integrates image representation network and classical image reconstruction method. The proposed model exhibits flexibility of choosing classical reconstruction method and powerful deep representation network. The application on MRI image reconstruction shows the effectiveness of the proposed method and it is also clear that the proposed method can be easily extended to other applications.

## References

1. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* **60**(1-4) (1992) 259–268
2. Cai, J.F., Dong, B., Osher, S., Shen, Z.: Image restoration: total variation, wavelet frames, and beyond. *Journal of the American Mathematical Society* **25**(4) (2012) 1033–1089
3. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: *Advances in neural information processing systems*. (2012) 341–349
4. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint* (2016)
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: *Advances in neural information processing systems*. (2014) 2672–2680
6. Sun, J., Li, H., Xu, Z., et al.: Deep admm-net for compressive sensing mri. In: *Advances in Neural Information Processing Systems*. (2016) 10–18
7. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine* **79**(6) (2018) 3055–3071
8. Adler, J., Öktem, O.: Learned primal-dual reconstruction. *arXiv preprint arXiv:1707.06474* (2017)
9. Adler, J., Öktem, O.: Solving ill-posed inverse problems using iterative deep neural networks. *arXiv preprint arXiv:1704.04058* (2017)
10. Mardani, M., Gong, E., Cheng, J.Y., Vasanawala, S., Zaharchuk, G., Alley, M., Thakur, N., Han, S., Dally, W., Pauly, J.M., et al.: Deep generative adversarial networks for compressed sensing automates mri. *arXiv preprint arXiv:1706.00051* (2017)
11. 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 transactions on medical imaging* **37**(2) (2018) 491–503
12. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. In: *Advances in neural information processing systems*. (2007) 153–160
13. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: *International Conference on Artificial Neural Networks*, Springer (2011) 52–59
14. Wang, R., Tao, D.: Non-local auto-encoder with collaborative stabilization for image restoration. *IEEE Transactions on Image Processing* **25**(5) (2016) 2117–2129
15. Fan, K., Wei, Q., Wang, W., Chakraborty, A., Heller, K.: Inversenet: Solving inverse problems with splitting networks. *arXiv preprint arXiv:1712.00202* (2017)
16. Fortin, M., Glowinski, R.: *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*. North-Holland Publishing Company (1983)
17. Bernstein, M.A., Fain, S.B., Riederer, S.J.: Effect of windowing and zero-filled reconstruction of mri data on spatial resolution and acquisition strategy. *Journal of Magnetic Resonance Imaging* **14**(3) (2001) 270–280
18. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine* **58**(6) (2007) 1182–1195