k-t NEXT: Dynamic MR Image Reconstruction Exploiting Spatio-Temporal Correlations

Chen Qin[1(✉)], Jo Schlemper[1], Jinming Duan[1,3], Gavin Seegoolam[1], Anthony Price[2], Joseph Hajnal[2], and Daniel Rueckert[1]

[1] Department of Computing, Imperial College London, London, UK
`c.qin15@imperial.ac.uk`
[2] Division of Imaging Sciences and Biomedical Engineering Department, King's College London, St. Thomas' Hospital, London, UK
[3] School of Computer Science, University of Birmingham, Birmingham, UK

Abstract. Dynamic magnetic resonance imaging (MRI) exhibits high correlations in k-space and time. In order to accelerate the dynamic MR imaging and to exploit k-t correlations from highly undersampled data, here we propose a novel deep learning based approach for dynamic MR image reconstruction, termed *k-t* NEXT (*k-t* NEtwork with X-*f* Transform). In particular, inspired by traditional methods such as *k-t* BLAST and *k-t* FOCUSS, we propose to reconstruct the true signals from aliased signals in x-f domain to exploit the spatio-temporal redundancies. Building on that, the proposed method then learns to recover the signals by alternating the reconstruction process between the x-f space and image space in an iterative fashion. This enables the network to effectively capture useful information and jointly exploit spatio-temporal correlations from both complementary domains. Experiments conducted on highly undersampled short-axis cardiac cine MRI scans demonstrate that our proposed method outperforms the current state-of-the-art dynamic MR reconstruction approaches both quantitatively and qualitatively.

1 Introduction

Dynamic Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique to monitor dynamic processes such as cardiac motion by acquiring data in a k-t space that contains both temporal and spatial information. However, the acquisition speed is limited due to both physical and physiological constraints. It is well known that in dynamic MRI there exists significant correlations in k-space and time. In order to increase the acquisition rate, most strategies have

been designed to acquire part of the desired k-t measurements and then reconstruct the images by exploiting spatio-temporal redundancies within the data.

Inspired by traditional k-t methods from the area of compressed sensing [8,9,15] for accelerated dynamic MR imaging, here we propose a novel dynamic MR image reconstruction NEtwork with X-f Transform, termed k-t NEXT, which exploits the signal redundancies in both x-f domain and image domain. In particular, the proposed k-t NEXT formulates the reconstruction process in an iterative fashion, where in each iteration, it consists of two sub-modules: a xf-CNN that learns to recover the true signals from aliased signals in x-f domain, and a convolutional recurrent neural network (CRNN) that exploits spatio-temporal redundancies in image domain. The dynamic reconstruction process thus alternates between x-f space and image space, which potentially enables the network to learn complementary features simultaneously from both domains. Experiments were performed on highly undersampled short-axis cardiac cine MR scans, where we show that the proposed model outperforms the current state-of-the-art dynamic MR reconstruction methods.

1.1 Related Work

Over the years, a number of approaches have been proposed for the reconstruction of accelerated dynamic MR images. In general, these methods can be mainly divided into three categories, based on exploiting correlations in k-space, in time, and in both k-space and time [15]. The first class of approaches exploit the correlations between k-space points at the same time frame, and then reconstruct each frame independently from other time frames, such as reduced field-of-view (FOV) [6] and parallel imaging methods [3], while the second group of strategies is to exploit redundancies in time, where the missing data at a given position can be interpolated or extrapolated from the measured data at other time points, such as keyhole imaging [7] and data sharing [18]. Relevant to our method, the third type of approaches is based on exploiting correlations in both k-space and time. One of the examples is the model-based k-t BLAST and k-t SENSE method [15], which takes advantage of a-priori information about the x-f support obtained from the training stage and then to remedy the aliasing artefacts during acquisition stage. Based on that, k-t FOCUSS [8,9] then formulated the problem in a compressed sensing MRI framework, which enforced the sparsity in x-f domain for the signal recovery. Similarly, a low rank and sparse reconstruction scheme (k-t SLR) [10] was proposed to exploit correlations between the temporal profiles of the voxels by introducing non-convex spectral norms and spatio-temporal total variation norm. In more recent years, deep learning approaches have gained their popularity for MR image reconstruction [2,12,13,16]. Most approaches investigate on exploiting information in a single frame (or static image) either in image domain [4,11,14] or in k-space domain [1,5,17], where each frame (or image) is reconstructed independently. In order to exploit the temporal redundancies, Schlemper et al. [13] proposed a data sharing (DS) layer in an image space cascaded 3D convolutional network to utilise the similar information contained in neighboring k-space samples. Qin et al. [12] also proposed a bidirectional CRNN model to exploit the temporal dependencies of dynamic sequences in

image domain. In contrast, our approach proposes to reconstruct the images in both x-f and image domains, where complementary information from two different domains can be fully exploited.

2 Methods

2.1 Problem Formulation

Consider a Cartesian k-space trajectory where k_x denotes the phase encoding direction, k_y denotes the readout direction, while $\sigma(x,t)$ denotes the image domain content at x and time t. The k-space measurement $v(k,t)$ is then formulated as:

$$v(k,t) = \int \sigma(x,t)e^{-j2\pi kx}dx = \int\int \rho(x,f)e^{-j2\pi(kx+ft)}dx\,df, \qquad (1)$$

where $\rho(x,f)$ is the 2D spectral signal in x-f domain. This can also be represented in a matrix form: $\mathbf{v} = \mathcal{F}\rho$, in which $\mathbf{v}$ and ρ stand for the stacked k-t space measurement vectors and x-f image respectively, and $\mathcal{F}$ is the 2D Fourier transform along the x-f direction. From the perspective of compressed sensing, the problem can be formulated by exploiting the sparsity of the unknown signal:

$$\min ||\rho||_1, \quad s.t. \ ||\mathbf{v} - \mathcal{F}\rho||_2 \leq \epsilon, \qquad (2)$$

where ϵ denotes the noise level. In k-t FOCUSS [8,9], the underdetermined inverse problem was solved via a sparse reconstruction algorithm called FOCUSS. The solution then can be expressed as the form that consists of a baseline signal $\bar{\rho}$ and its residual encoding for the n-th estimate of the x-f signal $\rho^{(n)}$:

$$\rho^{(n)} = \bar{\rho} + \text{FOCUSS}(\rho^{(n-1)} - \bar{\rho}, \rho^{(n-1)}). \qquad (3)$$

Here the mathematical form of FOCUSS algorithm is omitted for simplicity. For details, please refer to [8,9].

2.2 *k-t* NEXT for Dynamic MRI Reconstruction

Motivated by *k-t* BLAST [15] and *k-t* FOCUSS [9], we propose a dynamic image reconstruction NEtwork with X-*f* Transform (*k-t* NEXT) to exploit the spatio-temporal correlations from both x-f space and image space. Specifically, *k-t* NEXT formulates the iterative reconstruction process in an unfolded cascading way, as it has been shown to be a powerful technique in MR reconstruction [12,13]. In each iteration, our proposed approach learns to reconstruct the true images by alternating between x-f and image spaces, so that the spatio-temporal redundancies can be jointly exploited from these two complementary domains. In particular, a xf-CNN is proposed for the recovery of signals in x-f domain
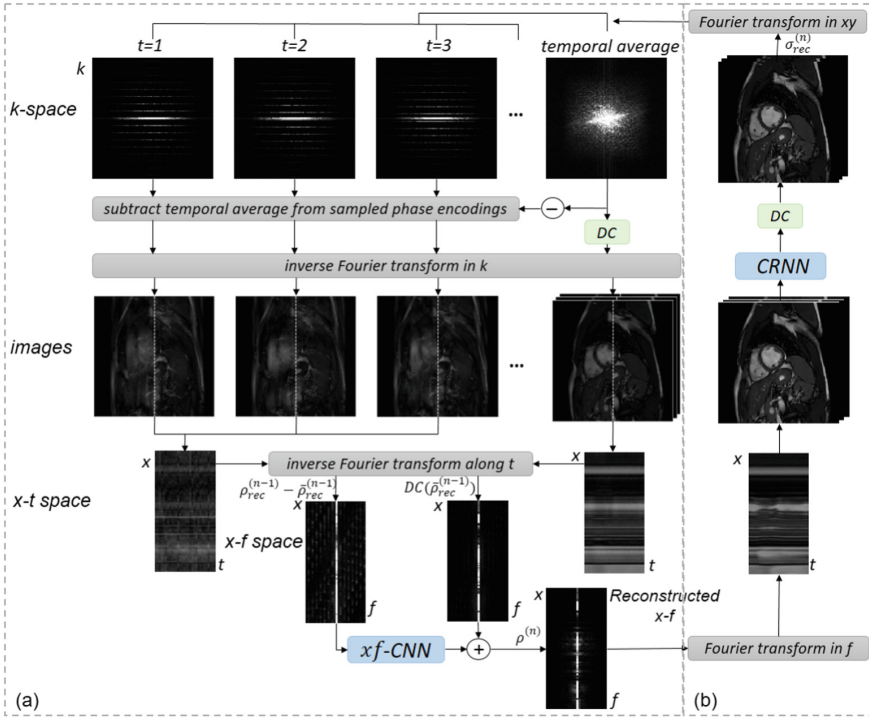
Fig. 1. The *k-t* NEXT reconstruction diagram. True signals can be recovered by iteratively updating the reconstruction in both (a) *x-f* and (b) image domains via learning the *xf*-CNN and CRNN jointly. For mathmetical notations, please refer to Eq. 4.

inspired by the traditional *k-t* method, and a variation of the CRNN-MRI [12] network is adopted for the subsequent image space reconstruction. We can compactly represent a single iteration of the *k-t* NEXT as follows:

$$\rho^{(n)} = \text{DC}(\bar{\rho}_{rec}^{(n-1)}) + xf\text{-CNN}(\rho_{rec}^{(n-1)} - \bar{\rho}_{rec}^{(n-1)}), \tag{4a}$$

$$\sigma_{rec}^{(n)} = \text{CRNN}(\mathcal{F}_f \rho^{(n)}; \mathbf{v}^{(0)}), \quad \rho_{rec}^{(n)} = \mathcal{F}_f^H \sigma_{rec}^{(n)}, \tag{4b}$$

where $\sigma_{rec}^{(n)} \in \mathbb{C}^D$ denotes the complex-valued reconstructed image sequence at iteration n, and $\sigma_{rec}^{(0)} = \sigma_u$ is the acquired zero-filled undersampled images. Here $D = D_x D_y T$, in which D_x and D_y are width and height of the frame and T is the number of frames. $\mathcal{F}_f$ denotes the Fourier transform along f dimension, and $\rho_{rec}^{(n)}$ is the *x-f* spectral signal transformed from $\sigma_{rec}^{(n)}$, while $\rho^{(n)}$ stands for the intermediate reconstructed signal from *xf*-CNN. Also $\bar{\rho}_{rec}^{(n-1)}$ denotes the temporally averaged *x-f* signal (see Eq. (5)), DC stands for the data consistency layer [13], and $\mathbf{v}^{(0)} \in \mathbb{C}^M$ ($M \ll D$) is the acquired raw data. An illustrative diagram of *k-t* NEXT is shown in Fig. 1. We will introduce it in the following.

xf-CNN Exploiting Spatio-Temporal Correlations in x-f Domain. Following the formulation in Eq. (3), here we propose to formulate the xf-CNN reconstruction as Eq. (4a), where instead of using model-based [15] or compressed sensing [9] algorithms to recover the true signals, we employ a stack of CNN layers to estimate the missing data based on other available points, typically within its vicinity in x-f space. In particular, here the x-f baseline signal $\bar{\rho}_{rec}^{(n)}$ is a temporal average of a sequence, i.e.,

$$\bar{\rho}_{rec}^{(n)} = \mathcal{F}^H \left[\sum_t \mathbf{v}^{(n)}./\max(\mathbf{1}, \sum_t \delta(\mathbf{v}^{(n)})) \right], \quad \delta(a) = \begin{cases} 0 & a = 0 \\ 1 & a \neq 0 \end{cases} \quad (5)$$

in which $\mathbf{v}^{(n)}$ is the k-space data that is Fourier transformed from $\sigma_{rec}^{(n)}$, and the ./ and max operation is performed element-wise. Thereby, xf-CNN learns to reconstruct residuals of each frame, which further exploits the signal sparsity.

The illustrative diagram of x-f reconstruction is shown in Fig. 1(a). Specifically, we formulate the k-t to x-f transformation process as a **x-f transform layer** in the network. In details, the x-f transform layer receives input from k-t space data. For iteration n, the acquired k-space data is firstly averaged along t to yield a temporal average (Eq. (5)), which is then subtracted from data at each time frame. To ensure data fidelity for the baseline estimate, here we propose to incorporate a data consistency (DC) term for $\bar{\rho}_{rec}^{(n-1)}$ at each frame separately. Then the subtracted data and temporally averaged data are inverse Fourier transformed to image space to obtain a sequence of aliased images and a data-consistent temporally averaged sequence. Each frequency-encoding position is then processed separately hereafter. The image columns from aliased images or baseline images are then gathered and inverse Fourier transformed along t to yield an x-f image, corresponding to $\rho_{rec}^{(n-1)} - \bar{\rho}_{rec}^{(n-1)}$ and DC($\bar{\rho}_{rec}^{(n-1)}$) respectively, which are then fed as inputs to xf-CNN for x-f space reconstruction (Eq. (4a)). After the signal de-aliasing in x-f domain, another Fourier transform along f is adopted to transform the estimated x-f signal $\rho^{(n)}$ back to dynamic image space for the subsequent image space reconstruction (Eq. (4b)).

k-t NEXT Exploiting Spatio-Temporal Redundancies in Complementary Domains. Previous approaches [2] have shown that exploring cross-domain knowledge is beneficial for MR reconstruction task. Inspired by this, with the aim of exploiting redundancies in complementary domains, here we propose to learn a dynamic MR reconstruction network in both x-f and image spaces jointly. In particular, we employ the CRNN model for image space reconstruction due to its effectiveness in exploiting temporal redundancies with a relatively smaller network capacity [12]. Thus, in each cascade, the proposed k-t NEXT consists of a xf-CNN and a CRNN block, where it employs all 2D convolutions across spatial and temporal dimensions, in contrast to 3D convolutions used in the baseline method [13]. This enables the network to be more efficient and effective in learning useful and complementary features in x-f, spatial and temporal space simultaneously.

Given the training data S with undersampled data as input and fully sampled data as target, i.e., (σ_u, σ_t) in image space and (ρ_u, ρ_t) in x-f space, the network is trained end-to-end by minimising the pixel-wise mean squared error (MSE) between the reconstructed data and the ground truth fully sampled data:

$$\mathcal{L}\left(\boldsymbol{\theta}\right) = \frac{1}{n_S} \sum \left(\left\| \sigma_t - \sigma_{rec}^{(N)} \right\|_2^2 + \left\| \rho_t - \rho^{(N)} \right\|_2^2 \right), \tag{6}$$

where $\sigma_{rec}^{(N)}$ and $\rho^{(N)}$ denote the predicted image and x-f array at iteration N, i.e., the final output in image domain and x-f domain respectively, $\boldsymbol{\theta}$ is the set of network parameters, and n_S is the number of training samples.

3   Experiments and Results

3.1   Dataset and Implementation Details

The dataset used in our experiments consists of 10 fully sampled complex-valued short-axis cardiac cine MRI. Each scan contains a single slice SSFP acquisition with 30 temporal frames. The raw data has 32-channel data with sampling matrix 192×190, which was zero-filled to 256×256, and the raw multi-coil data was then reconstructed to produce a single complex-valued image. In experiments, images were transformed back to k-space to simulate a fully sampled single-coil acquisition. A shear grid k-t Cartesian sampling pattern with four central lines (see Fig. 3(b)) was employed to undersample the k-space data to generate the undersampled input image sequences. The undersampling rate mentioned is stated with respect to the matrix size of the data, which is 192×190.

In the proposed k-t NEXT, xf-CNN is composed of 5 layers of 2D CNN with a residual connection from the baseline estimate. For the CRNN model, a variation of architecture [12] is employed which consists of 4 layers of bidirectional CRNN, 1 layer of 2D CNN, a residual connection and a DC layer. We used dilated convolutions with kernel size 3×3 and dilation factor $(3, 3)$, and the number of cascade N was set to 4 for all comparison methods. For detailed network architecture, please refer to supplementary materials. The network was implemented in PyTorch. During training, ADAM optimiser was employed with a learning rate of 10^{-4}. Data augmentation was performed on-the-fly, with random rotation, scaling, and elastic transformation. All evaluations were done via a 3-fold cross validation.

3.2   Results

In experiments, we compared our proposed approach (k-t NEXT) with different dynamic MR reconstruction methods, including compressed sensing method k-t FOCUSS [9], deep learning method CRNN-MRI [12], and DS+3DCNN [13] that incorporates data sharing (DS). To investigate the effectiveness of xf-CNN, an additional baseline approach is proposed which replaces all x-f reconstruction in k-t NEXT with DS component, termed DS+CRNN. In DS methods, we set

Table 1. Comparison results of different methods on dynamic cardiac cine MRI with high undersampling rate 9 and 12. Best results are indicated in bold.

| Method		*k-t* FOCUSS	CRNN-MRI	DS+3DCNN	DS+CRNN	*k-t* NEXT
Capacity		-	260,866	352,770	265,474	374,020
9×	PSNR	29.52 (1.58)	32.45 (1.33)	33.47 (1.41)	33.24 (1.38)	**34.23** (1.44)
	SSIM	0.951 (0.013)	0.969 (0.008)	0.975 (0.006)	0.975 (0.006)	**0.979** (0.005)
	HFEN	0.340 (0.033)	0.249 (0.032)	0.214 (0.026)	0.215 (0.027)	**0.196** (0.030)
12×	PSNR	28.14 (1.56)	31.30 (1.32)	32.46 (1.36)	32.34 (1.35)	**33.18** (1.40)
	SSIM	0.937 (0.016)	0.962 (0.009)	0.969 (0.007)	0.970 (0.007)	**0.975** (0.005)
	HFEN	0.382 (0.035)	0.282 (0.034)	0.242 (0.027)	0.239 (0.029)	**0.225** (0.031)
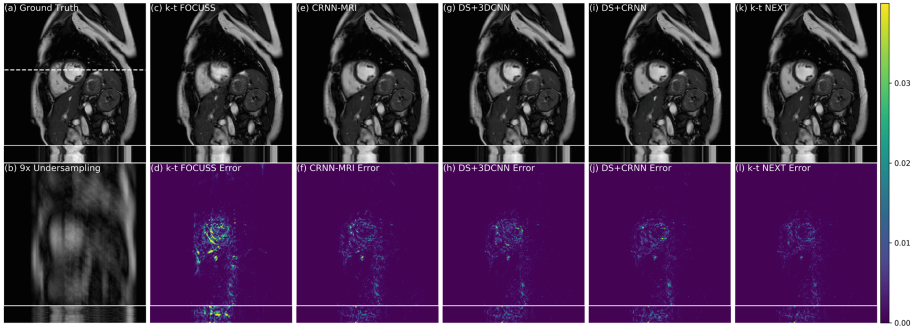
Fig. 2. Comparison results on spatial and temporal dimensions with their error maps. A dynamic video is shown in supplementary materials for better visualisation.

Fig. 3. Visualisation in *x-f* domain. (a) Ground Truth (b) *k-t* sampling pattern (c) 9× undersampled data (d) Reconstructed *x-f* image (e) Error between (c) and (d).

the number of neighbouring frame as $n_{adj} \in \{0, 1, ...5\}$ as in [13]. Note that for a fair comparison with our *k-t* NEXT, we modified the baseline approaches DS+3DCNN and DS+CRNN to learn the residual of a temporally averaged frame as well. Quantitative comparison results of different methods on dynamic cardiac data with undersampling rates 9 and 12 are presented in Table 1, where it compares the network capacity per cascade, peak-to-noise-ratio (PSNR), structural similarity index (SSIM) and high frequency error norm (HFEN) [12]. Networks for different undersampling factors were trained separately in this case. It can be seen that our proposed *k-t* NEXT can outperform other baseline methods by a large margin in terms of all these measures at different undersampling rates, with roughly the same level of network capacity. In particular, *k-t* NEXT performs better than its corresponding DS pair, which indicates the merits of exploiting correlations in *x-f* space and complementary domains.

Additionally, we compared the qualitative results on $9\times$ undersampled data in Fig. 2, where it shows the reconstructed images along both spatial and temporal dimensions, as well as their corresponding error maps. It can be observed that our proposed model can faithfully recover the images with smaller errors especially around dynamic regions compared with other baseline methods. In particular, k-t NEXT produced visually sharper images than DS methods. This is reflected by the fact that, in contrast to DS approaches which fill in k-space data from neighboring frames and therefore could possibly generate averaged and smooth images, k-t NEXT directly estimates the missing data in x-f space. A visualisation of x-f reconstruction is also presented in Fig. 3, where it displays the reconstructed x-f image and its error map in comparison to the input aliased data. It can be observed that the aliasing artefacts were largely removed and the undersampled data were recovered to approximate the ground truth signals.

4   Conclusion

In this paper, we have presented a novel deep learning based method, k-t NEXT (k-t NEtwork with X-f Transform), for highly undersampled dynamic MR image reconstruction. xf-CNN is proposed to exploit correlations in k-t space via reconstructing the true signals from aliased signals in x-f domain. Based on that, k-t NEXT is then proposed to learn to iteratively recover the images by alternating between the complementary x-f and image domains, where networks from both domains were trained jointly. Experimental results have shown that the proposed k-t NEXT outperforms state-of-the-art dynamic MR reconstruction methods in terms of both quantitative and qualitative performance. For the future work, we will extend the method for dynamic 3D applications.

References

1. Akçakaya, M., Moeller, S., Weingärtner, S., et al.: Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. Magn. Reson. Med. **81**(1), 439–453 (2019)
2. Eo, T., Jun, Y., Kim, T., et al.: KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magn. Reson. Med. **80**(5), 2188–2201 (2018)
3. Griswold, M.A., Jakob, P.M., Heidemann, R.M., et al.: Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med. **47**(6), 1202–1210 (2002)
4. Hammernik, K., Klatzer, T., Kobler, E., et al.: Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med. **79**(6), 3055–3071 (2018)
5. Han, Y., Ye, J.C.: k-space deep learning for accelerated MRI. arXiv preprint arXiv:1805.03779 (2018)

6. Hu, X., Parrish, T.: Reduction of field of view for dynamic imaging. Magn. Reson. Med. **31**(6), 691–694 (1994)
7. Jones, R., Haraldseth, O., Müller, T., et al.: K-space substitution: a novel dynamic imaging technique. Magn. Reson. Med. **29**(6), 830–834 (1993)
8. Jung, H., Sung, K., Nayak, K.S., et al.: k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI. Magn. Reson. Med. **61**(1), 103–116 (2009)
9. Jung, H., Ye, J.C., Kim, E.Y.: Improved k-t BLAST and k-t SENSE using FOCUSS. Phys. Med. Biol. **52**(11), 3201 (2007)
10. Lingala, S.G., Hu, Y., DiBella, E., et al.: Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. IEEE Trans. Med. Imaging **30**(5), 1042–1054 (2011)
11. Liu, J., Kuang, T., Zhang, X.: Image reconstruction by splitting deep learning regularization from iterative inversion. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 224–231. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_26
12. Qin, C., Schlemper, J., Caballero, J., et al.: Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging **38**(1), 280–290 (2019)
13. Schlemper, J., Caballero, J., Hajnal, J.V., et al.: A deep cascade of convolutional neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging **37**(2), 491–503 (2018)
14. Schlemper, J., et al.: Bayesian deep learning for accelerated MR image reconstruction. In: Knoll, F., Maier, A., Rueckert, D. (eds.) MLMIR 2018. LNCS, vol. 11074, pp. 64–71. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00129-2_8
15. Tsao, J., Boesiger, P., Pruessmann, K.P.: k-t BLAST and k-t SENSE: dynamic MRI with high frame rate exploiting spatiotemporal correlations. Magn. Reson. Med. **50**(5), 1031–1042 (2003)
16. Ye, J.C., Han, Y., Cha, E.: Deep convolutional framelets: a general deep learning framework for inverse problems. SIAM J. Imaging Sci. **11**(2), 991–1048 (2018)
17. Zhang, P., Wang, F., Xu, W., et al.: Multi-channel generative adversarial network for parallel magnetic resonance image reconstruction in k-space. In: MICCAI, pp. 180–188 (2018)
18. Zhang, S., Block, K.T., Frahm, J.: Magnetic resonance imaging in real time: advances using radial FLASH. Magn. Reson. Med. **31**(1), 101–109 (2010)