Journal of Magnetic Resonance 305 (2019) 232–246

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Undersampled MR image reconstruction using an enhanced recursive residual network

Lijun Bao [a,*], Fuze Ye [a], Congbo Cai [a], Jian Wu [a], Kun Zeng [a], Peter C.M. van Zijl [b,c], Zhong Chen [a]

[a] Department of Electronic Science, Xiamen University, Xiamen 361000, China
[b] Department of Radiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[c] F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD 21205, USA

ARTICLE INFO

Article history:
Received 27 March 2019
Revised 24 June 2019
Accepted 8 July 2019
Available online 9 July 2019

Keywords:
Convolutional neural network
Error-correction
Feature guidance
Recursive residual learning
Undersampled MRI reconstruction

ABSTRACT

When using aggressive undersampling, it is difficult to recover the high quality image with reliably fine features. In this paper, we propose an enhanced recursive residual network (ERRN) that improves the basic recursive residual network with a high-frequency feature guidance, an error-correction unit and dense connections. The feature guidance is designed to predict the underlying anatomy based on image a priori learned from the label data, playing a complementary role to the residual learning. The ERRN is adapted for two important applications: compressed sensing (CS) MRI and super resolution (SR) MRI, while an application-specific error-correction unit is added into the framework, i.e. data consistency for CS-MRI and back projection for SR-MRI due to their different sampling schemes. Our proposed network was evaluated using a real-valued brain dataset, a complex-valued knee dataset, pathological brain data and *in vivo* rat brain data with different undersampling masks and rates. Experimental results demonstrated that ERRN presented superior reconstructions at all cases with distinctly restored structural features and highest image quality metrics compared to both the state-of-the-art convolutional neural networks and the conventional optimization-based methods, particularly for the undersampling rate over 5-fold. Thus, an excellent framework design can endow the network with a flexible architecture, fewer parameters, outstanding performances for various undersampling schemes, and reduced overfitting in generalization, which will facilitate real-time reconstruction on MRI scanners.

1. Introduction

Magnetic resonance imaging (MRI) is a non-invasive diagnostic technique providing computerized images of internal body tissues. Because of its excellent contrast in soft tissues, MRI is used widely to assess brain diseases, cardiac function, angiography and so on. However, its long acquisition time causes discomfort to patients and thus hinders the time-critical applications. The image quality in MRI scanning may also be corrupted by various artifacts and subject movements. In this case, compressed sensing (CS) and parallel imaging are commonly utilized to accelerate MRI. CS approach exploits image sparsity to reconstruct high quality images from the undersampled k-space data [1], so as to reduce the acquisition time. On the other side, super resolution (SR) technique is used as a post processing to assist the fast and ultrafast imaging [2], e.g. in cardiac and fetal MRI [3]. Although zero-filling in k-space for CS-MRI or Bicubic interpolation in the spatial domain for SR-MRI can improve the apparent image resolution in easy ways, they introduce no additional information in terms of contrast. The zero-filling inevitably results in low quality observations with aliasing artifacts proportional to the acceleration rate. Therefore, how to recover the high quality image with reliably tissue structures and increased SNR from undersampled measurements is a difficult problem in undersampled MR image reconstruction.

In the past few decades, many optimization-based algorithms have been proposed combining with various regularization constraints, e.g. total variation (TV) [4], sparse a priori in a certain transform domain or a dictionary [5,6]. These methods perform better than interpolation methods, but their reconstructed images will deteriorate at aggressive undersampling rates, especially over 5-fold. In optimization-based methods, a nonlinear reconstruction algorithm should balance the consistency with acquired data in k-space against the sparsity in transform domain. In addition, it is a challenging issue to determine the appropriate regularization and the transform domain, and these methods generally suffer from time-consuming iterations, thus restricting their clinical

---

[*] Corresponding author at: Department of Electronic Science, Xiamen University, Xiamen 361005, China.

E-mail address: baolijun@xmu.edu.cn (L. Bao).

https://doi.org/10.1016/j.jmr.2019.07.020
1090-7807/

applications. Recently, the development of modern GPUs has made it feasible to use deep convolutional neural network (CNN) in the computer processing. CNN can integrate millions of parameters and multiple layers into a complicated network architecture, which has attained the state-of-the-art performance in many applications, such as object classification [7] and recognition [8,9], image segmentation [10] and restoration [11]. In particular, deep CNNs have been developed for image reconstructions and obtained outstanding performances [12]. CNN was initially adopted to learn an end-to-end mapping between low quality inputs and labels with 3 convolutional layers [13]. Subsequently, a deep framework of 20 convolutional layers has been designed to exploit the contextual information over larger receptive field [14]. In the following, researchers proposed a recursive CNN to deepen the network without increasing the parameter number [15]. Meanwhile, residual learning was explored to facilitate the network training [8] and utilized in generator adversarial nets (GAN) [16].

Motivated by the great success in natural images, CNN-based methods become an interesting topic in the biomedical imaging due to its significant potential for clinical application [17]. Through some modifications in aspects of loss function, network input, convolution dimension, anatomy priors and so on, existing CNN frameworks were applied to single- or multi-modality MRI reconstructions. At early stage, CNN-based CS-MRI directly learned the mapping between zero-filling inputs and the corresponding fully-sampled data [18]. Afterwards, a deep ADMM-net reformulated the algorithm of alternating direction method of multipliers (ADMM) for an increased computation speed [19]. To allow fast and high quality reconstruction of clinical accelerated multi-coil MRI data, a variational network was proposed that combined the mathematical structure of variational models with deep learning [20]. Meanwhile, data consistency has been exploited in a deep cascade CNN framework, referred as DC-CNN hereinafter [21]. In addition, the well-known U-net is a popular architecture in medical image reconstruction problems [22,23,24]. GAN employing the deep generator and discriminator networks [25] also presented outstanding performance in CS-MRI [26,27,28]. Moreover, the residual learning was used in SR-MRI reconstruction [29,30,31], and a regularized network incorporating anatomical prior has been proposed for the cardiac image enhancement [32]. The attractive AUTOMAP (automated transform by manifold approximation) succeeded in learning reconstruction transforms for various MRI acquisition strategies, including both Cartesian and non-Cartesian samplings, using the same network architecture and hyperparameters, nevertheless the training demands huge memory allocation [33].

In this paper, we aim to improve general performance by introducing physically meaningful components to the network. Differing from networks designed for a single task, we also intend to enable a network framework to be adapted for two important applications: CS-MRI and SR-MRI. Here, we propose an enhanced recursive residual network (ERRN) for undersampled MR image reconstruction in CS and SR. Our ERRN framework is based on a recursive residual network and enhanced with a set of user-designed functional modules, i.e. a high-frequency feature guidance, an error-correction unit and improved dense connections. The feature guidance is designed to emphasize underlying anatomical structures using predefined convolutional kernels. As for the error-correction unit, data consistency in k-space is adapted to fulfill data fidelity for CS-MRI, whereas back projection in spatial domain is utilized for SR-MRI. In addition, dense connections encourage features propagating to the latter layers in the deep network. ERRN was evaluated using a real-valued brain dataset, a complex-valued knee dataset, clinical brain tumor data and *in vivo* rat brain data. It achieved superior performances on different CS sampling patterns at rates 10%, 20% and 30%, as well as for SR at sampling scales 2×2, 3×3 and 4×4. Compared to other normal networks, the success of ERRN has demonstrated that a sophisticated architecture enhanced by functional modules is more powerful and adaptive for diverse applications, while fewer network parameters also can reduce overfitting in the generalization.

2. Theory

In this section, the mathematical model of undersampled MRI reconstruction is introduced and then the network loss function is derived.

2.1. Mathematical model

The MRI undersampling acquisition is generally formulated as $\mathbf{Y} = \mathbf{F}_u\mathbf{X}$, where $\mathbf{X} \in \mathbb{C}^M$ is the fully-sampled reference image of size $\sqrt{M} \times \sqrt{M}$, and $\mathbf{Y} \in \mathbb{C}^M$ is the undersampled observation in k-space where missing data is padded by zero. The linear operator $\mathbf{F}_u$ is defined as $\mathbf{F}_u = \mathbf{DF}$, where $\mathbf{F}$ denotes the Fourier transform and $\mathbf{D}$ is a binary mask denoting the k-space undersampling scheme where the sampled index is indicated by one. In spite of different undersampling schemes, MRI reconstruction aims to recover high quality images with accurate and faithful anatomical structures from undersampled measurements. In optimization-based methods, reconstructed image $\widehat{\mathbf{X}}$ can be solved using the mathematical model as

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \, ||\mathbf{F}_u\mathbf{X} - \mathbf{Y}||_2^2 + \gamma \cdot \Phi(\mathbf{X}) \tag{1}$$

where the first term is to calculate the data fidelity and the second term is a regularization constraint weighted by parameter γ. The constraint may be based on the intensity information, prediction models, edge-based or patch-based operators, e.g. total variation [1], sparse coding [6], non-local similarity [34,35], nearest neighbors or other constraints.

2.2. Network loss function

However, CNN is capable to integrate feature learning, nonlinear mapping and image reconstruction into a direct end-to-end network. To determine the mapping function from $\mathbf{Y}$ to $\mathbf{X}$, it is essential to learn the optimal value of network parameters denoted as $\mathbf{\Theta}$, which can be implemented by minimizing a loss function in network training. Given the training dataset containing $\{\mathbf{Y}_t\}$ and $\{\mathbf{X}_t\}$, with $t = 1, \cdots, \mathrm{T}$, denoting the sample number, the loss function is defined as an optimization objective

$$\mathrm{L}(\Theta) = \sum_{t=1}^{T} ||\mathbf{F}_u\hat{\mathbf{X}}_t - \mathbf{Y}_t||_2 + \gamma_1 \cdot \Phi\left(\hat{\mathbf{X}}_t\right) + \gamma_2 \cdot ||\hat{\mathbf{X}}_t - \mathbf{X}_t||_2 \tag{2}$$

where the first term is to measure data fidelity and the second one means a regularization constraint, while the last plays a role of error-correction. Once the mapping relationship is learned from the network training, network parameters $\mathbf{\Theta}$ are determined, and then we can attain the reconstructed result $\widehat{\mathbf{X}}$ corresponding to the network input $\mathbf{Y}$. How to represent these terms using convolutional layers in a sophisticated network is our main research interest in this work [36,37].

3. Methods

In this section, the basic architecture of our recursive residual network is described firstly. Then, we propose the enhanced recursive residual network (ERRN) to solve undersampled image reconstructions in CS-MRI and SR-MRI, in which the basic network will be enhanced by a high-frequency feature guidance, an error-correction unit and improved dense connections. Particularly, the error-correction unit is adapted for different reconstruction tasks, i.e. data consistency for CS-MRI and back projection for SR-MRI.

3.1. Basic recursive residual network

3.1.1. Network architecture

The recursive residual network used for image reconstruction is motivated by the work in [37] and [38]. The basic architecture of our recursive residual network is illustrated in Fig. 1. It consists of three subnetworks: an embedding net, an inference net and a reconstruction net [15]. The embedding net is created to extract structural features from network inputs, in which the zero-filling image is denoted as $\mathbf{X}_u = \mathbf{F}^{\mathrm{H}}\mathbf{Y}$, where $\mathbf{F}^{\mathrm{H}}$ is the inverse Fourier transform. The network processes complex-valued data with two separate channels for real and imaginary components respectively, or just single channel for real-valued one. The inference net is stacked by a set of parameter-shared residual blocks where training is executed in a multi-supervision strategy. In this way, feature maps generated by each residual block are convolved and then summed up with the feedforward in the reconstruction net. Consequently, the intermediate prediction $\widehat{\mathbf{X}}_i$ of each block is involved in the final reconstruction defined by $\widehat{\mathbf{X}} = \sum_{i=1}^{n}\omega_i\widehat{\mathbf{X}}_i$, where ω_i is the averaging weight and n is the number of residual blocks. The recursive residual block is rendered in light blue and denoted with RB*i* in Fig. 1c. For the sake of concision, we format the weighted averaging part as *Multi-Supervision* and will omit its detail in the following ERRN architectures.

3.1.2. Recursive residual blocks

Rather than learning unreferenced functions, a residual block learns the residual function with reference to the layer inputs, which can effectively mitigate the training difficulty in deep networks [8]. In general, deep networks with more convolutional layers may have better performance, but the substantially increased parameters require more memory. Therefore, we explore the recursive learning [15,39] to construct a deep inference net, which is consistent with the iterative operation in optimization-based methods. In this case, all residual blocks share the same parameters and thus more residual blocks will not cause additional parameters. More elucidation of the relationship between recursive residual network and sub-band based image reconstruction method [38] is provided in supplementary material with Fig. S1.

Our residual block is composed of two pre-activation convolutional layers and one skip connection between the input and the output, as depicted in Fig. 1b. The pre-activation performs activation functions before the convolutional layer, including batch normalization (BN) [40] and rectified linear unit (ReLU*)* [41] as depicted in Fig. 1a, which could well preserve the negative information for the reconstruction task [39]. Denoting the input and output of *i* th residual block as $\mathbf{B}_{i-1}$ and $\mathbf{B}_i$, the residual block can be formulated as

$$\mathbf{B}_i = f_{RB}(\mathbf{B}_{i-1}) + \mathbf{B}_{i-1} = \mathbf{W}_{RB}^2\left(\tau\left(\mathbf{W}_{RB}^1(\tau(\mathbf{B}_{i-1}))\right)\right) + \mathbf{B}_{i-1},$$
$$i = 1, 2, \cdots, n \tag{3}$$

where $f_{RB}(\cdot)$ is the residual function, $\mathbf{W}_{RB}^1$ and $\mathbf{W}_{RB}^2$ represent the weights, and τ denotes activation functions. The bias term is omitted here for simplicity.

3.2. ERRN for CS-MRI

3.2.1. High-frequency feature guidance

High-frequency information in *k*-space often contains important structures that are prone to be corrupted but hard to recover. We address this by putting more effort to reconstruct those details, i.e. adding a functional module of feature guidance into the reconstruction net, which is a constraint penalizing the deviation from guiding features.

Four filters along horizontal, vertical and diagonal directions are created to extract high-frequency features as shown in Fig. 2. For the zero-filling image, feature maps are detected by $\mathbf{X}_u^h = \mathbf{G}\otimes\mathbf{X}_u$ with $\mathbf{G}$ denoting the directional filters, while guiding features are extracted from the reference image as $\mathbf{X}^h = \mathbf{G}\otimes\mathbf{X}$. The low quality images and their high-frequency features are concatenated as network inputs in the form of $[\mathbf{X}_u, \mathbf{X}_u^h]$. We extract high-frequency features from the real and imaginary parts, respectively. In the ERRN framework shown in Fig. 3, feature guidance module is outlined by a dashed border, where feature maps $\mathbf{F}_{r1i}$ flow into the *FeaRecon* layer to learn the underlying anatomy under the guidance of $\mathbf{X}^h$.

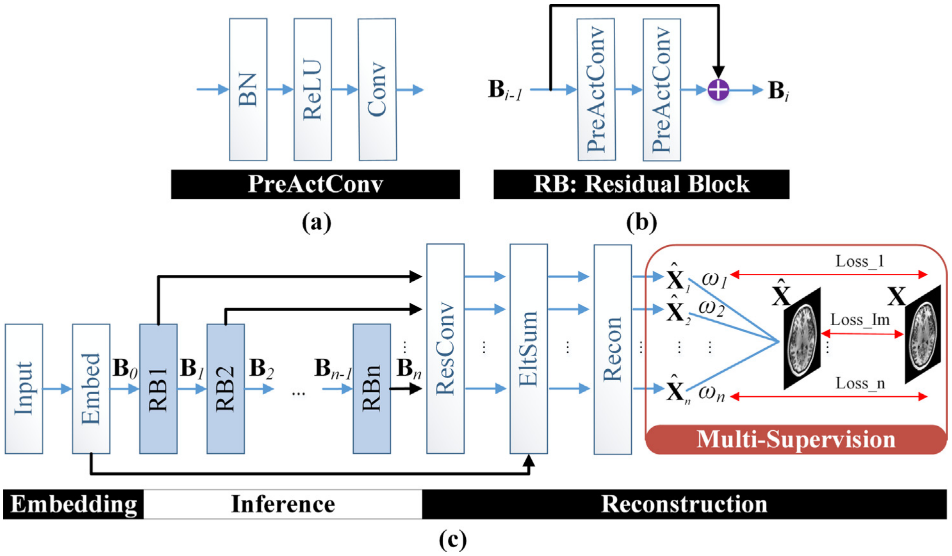

Fig. 1. Basic recursive residual network. The constitution of a residual block is depicted in (a) and (b). *Embed*, *Resconv* and *Recon* in (c) are pre-activation convolutional layers. *PreActConv*: pre-activation convolutional layer. *EltSum*: elementwise summation.

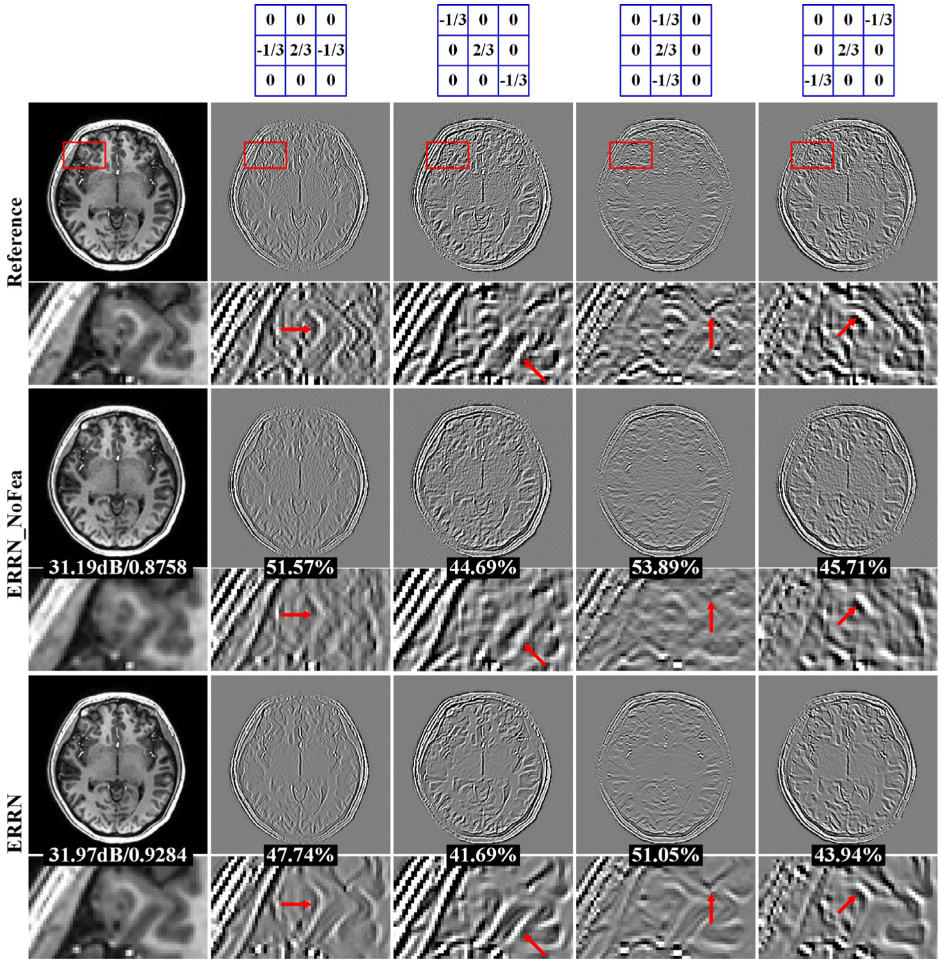

Fig. 2. Directional filters **G** and their corresponding high-frequency feature maps extracted from different images. Reference: fully-sampled image, ERRN_Nofea: ERRN framework without feature guidance. ERRN: our proposed network. The feature maps are $\widehat{\mathbf{X}}_{10}^h$, learned from the 10th residual block, and their NRMSE values are relative to $\mathbf{X}^h$, noted in percent format. PSNR/MSSIM are provided with reconstructed images.

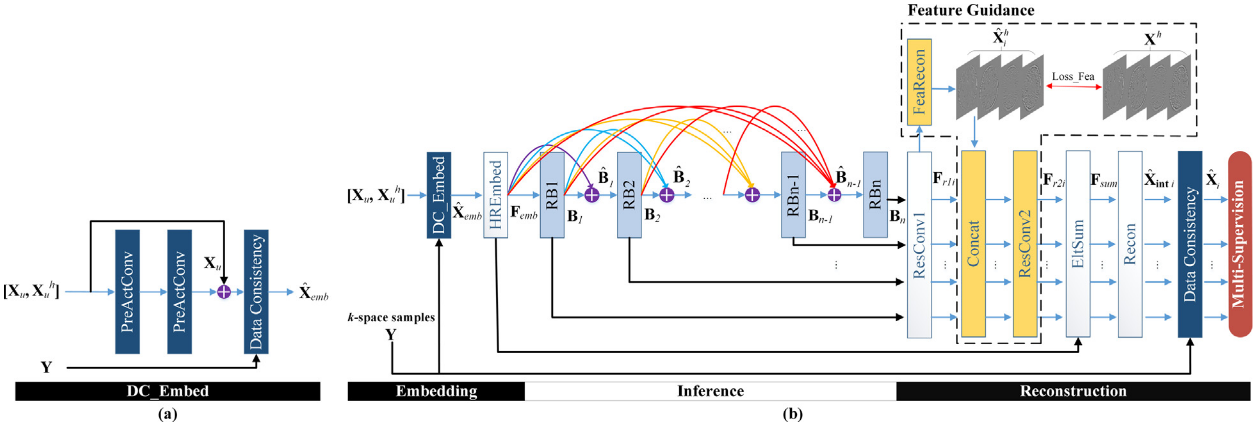

Fig. 3. ERRN network architecture for CS-MRI. (a) *DC_Embed* block, (b) ERRN framework. Data consistency is adopted in embedding and reconstruction nets as noted in dark blue. The dashed border outlines the high-frequency feature guidance in yellow blocks. Colored curves denote dense connections. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The learned features $\widehat{\mathbf{X}}_i^h$ are fed back to the main framework and are concatenated into $\left[\mathbf{F}_{r1i}, \widehat{\mathbf{X}}_i^h\right]$.

In Fig. 2, the effect of feature guidance is demonstrated in CS-MRI using the variable-density random sampling at rate 10% [42], while the corresponding demonstration in SR-MRI is arranged in our supplementary material with Fig. S2. Taking $\mathbf{X}^h$ as reference, tissue structures recovered by ERRN are closer to the ground truth than those of ERRN_NoFea (without feature guidance), as noted by red arrows. Quantitatively, feature maps of the ERRN result present smaller NRMSE (normalized root mean squared error).

PSNR/MSSIM values of the reconstructed images further indicate that ERRN outperforms ERRN_NoFea obviously, with 31.97 dB/0.9284 verse 31.19 dB/0.8758.

3.2.2. Improved dense connections

As the network depth growing, the hierarchical features of previous layers would get lost in the latter, while the problem of vanishing gradients will hamper the network training. To address this problem, ResNet [8] and U-net [10] use the skip connection to create a shortcut connection between any two layers. As the extreme case of skip connection, the dense connection connects each layer to all forward layers, which can greatly alleviate the gradients vanishing problem and strengthen the feature propagation [9]. Dense connections have been utilized by means of using 1 × 1 convolutional layer as a gate unit to learn block weights [12] or concatenating feature maps of preceding layers with parameters growing exponentially [43]. Here, we introduce dense connections to the inference net as delineated by colorful curves in Fig. 3. This helps ERRN to maintain propagation of the output features of every residual block, and thus facilitates the gradients flow and the complicated structures learning. It is worth mentioning that, considering the recursive learning strategy, our dense connections only assign a trainable scalar weight μ_j to feature maps of the previous block $\mathbf{B}_j$, and the weighted average is fed forward to the current block. Hence, the input of each residual block $\widehat{\mathbf{B}}_i$ is rewritten as

$$\hat{\mathbf{B}}_i = \sum_{j=0}^{i} \mu_j \mathbf{B}_j, j = 0, 1, \cdots i, \ \ i = 1, 2, \cdots n-1 \tag{4}$$

3.2.3. Data consistency in ERRN

CS-MRI has been successfully applied to scan acceleration and image enhancement, which relies on the undersampling acquisition protocol by Cartesian [1] or non-Cartesian [44] sampling trajectories to achieve incoherent artifacts in a certain transform domain. So far, various undersampling schemes have been proposed for MRI acquisition. In this paper, we focus on non-Cartesian sampling which is available for high acceleration rates over × 5. Since the system bias would accumulate when signals flow across the deep network, we introduce a functional module of error-correction into ERRN framework. In CS-MRI application, we employ data consistency to design the error-correction unit. Data consistency is adopted in embedding and reconstruction nets, as noted by dark blue in Fig. 3. Given $\widehat{\mathbf{X}}_{\mathrm{int}\_i}$ as the intermediate prediction of each residual block, its corresponding value in *k*-space can be expressed as $\widehat{\mathbf{Y}}_{\mathrm{int}\_i} = \mathbf{F}\widehat{\mathbf{X}}_{\mathrm{int}\_i}$. Then, the output of *Data Consistency* layer is written as

$$\hat{\mathbf{X}}_i = f_{\mathrm{DC}}\left(\hat{\mathbf{X}}_{\mathrm{int}\_i}, \mathbf{Y}\right) = \mathbf{F}^{\mathrm{H}} \frac{\hat{\mathbf{Y}}_{\mathrm{int}\_i}(\mathbf{k}) + \lambda \mathbf{Y}(\mathbf{k})}{1 + \lambda} \tag{5}$$

where $\mathbf{k}$ represents the *k*-space index, and λ is a positive constant. Thereby, ERRN implements a linear fitting between network predictions and original measurements, taking λ to adjust the weight. The parameter λ is related to imaging noise level, i.e. smaller value for stronger noise, and it is determined in training.

3.2.4. Framework implementation

The network architecture of ERRN for CS-MRI is depicted in Fig. 3 and its implementation is outlined in Algorithm 1. We build our training objective function based on a linear loss combination of the high-frequency feature guidance $\widehat{\mathbf{X}}_i^{\boldsymbol{h}}$, the prediction of each residual block $\widehat{\mathbf{X}}_i$ and the final reconstruction $\widehat{\mathbf{X}}$. The loss function of ERRN is formulated into

$$\mathrm{L}(\Theta) = \sum\nolimits_{t=1}^{\mathrm{T}} \| \mathbf{X}^t - \sum\nolimits_{i=1}^{n} \omega_i \hat{\mathbf{X}}_i^t \|_2 + \sum\nolimits_{t=1}^{\mathrm{T}} \sum\nolimits_{i=1}^{n} \| \mathbf{X}^t - \hat{\mathbf{X}}_i^t \|_2 + \sum\nolimits_{t=1}^{\mathrm{T}} \sum\nolimits_{i=1}^{n} \| \mathbf{X}^{ht} - \hat{\mathbf{X}}_i^{ht} \|_2 \tag{6}$$

where T denotes the number of training samples and *n* is the number of residual blocks. Our network is trained based on the optimization approach in [45].

Algorithm 1. ERRN for CS-MRI Reconstruction

```
1: Input $\mathbf{Y}$, $\mathbf{X}_{\mathrm{u}}$ and $\mathbf{X}_{\mathrm{u}}^{\mathrm{h}}$, setting $\mu_{\mathrm{j}}$ in (4) as 1/j and the weight $\omega_{\mathrm{i}}$
   as 1/n, other parameters are initialized following MSRA in
   [46].
Embedding net
2: Extract features from the input data with DC_Embed and
   HREmbed
$\widehat{\mathbf{X}}_{emb} = f_{DC}\left(f_{emb}\left(\left[\mathbf{X}_u, \mathbf{X}_u^h\right]\right), \mathbf{Y}\right), \mathbf{F}_{emb} = \mathbf{W}_{emb}\left(\tau\left(\widehat{\mathbf{X}}_{emb}\right)\right)$
Inference net
3: Calculate input and output of each residual block using (3)
   and (4).
Reconstruction net
4: Feature maps generated from $\mathbf{B}_{\mathrm{i}}$ are convolved in ResConv1:
$\mathbf{F}_{r1i} = f_{r1}(\mathbf{B}_i) = \boldsymbol{W}_{r1}(\tau(\mathbf{B}_i))$
5: Learn the underlying anatomy using the feature guidance
   FeaRecon:
$\widehat{\mathbf{X}}_i^h = f_{frc}(\mathbf{F}_{r1i}) = \mathbf{W}_{frc}(\tau(\mathbf{F}_{r1i}))$
6: Concatenation $\left[\mathbf{F}_{\mathrm{r1i}}, \widehat{\mathbf{X}}_{\mathrm{i}}^{\mathrm{h}}\right]$ is convolved in ResConv2:
$\mathbf{F}_{r2i} = f_{r2}\left(\left[\mathbf{F}_{r1i}, \widehat{\mathbf{X}}_i^h\right]\right) = \mathbf{W}_{r2}\left(\tau\left(\left[\mathbf{F}_{r1i}, \widehat{\mathbf{X}}_i^h\right]\right)\right)$
7: Connect with embedding net by a skip connection EltSum:
$\mathbf{F}_{sum} = \mathbf{F}_{emb} + \mathbf{F}_{r2i}$
8: The intermediate prediction of ith residual block Recon:
$\widehat{\mathbf{X}}_{\mathrm{int}\_i} = f_{rc}(\mathbf{F}_{sum}) = \mathbf{W}_{rc}(\tau(\mathbf{F}_{sum}))$
9: Update the prediction with Data Consistency:
$\widehat{\mathbf{X}}_i = f_{DC}\left(\widehat{\mathbf{X}}_{\mathrm{int}\_i}, \mathbf{Y}\right)$
10: Final reconstruction using Multi-Supervision:
   $\widehat{\mathbf{X}} = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \omega_{\boldsymbol{i}} \widehat{\mathbf{X}}_i$
```

3.3. ERRN extending to SR-MRI

Independent from any sequence or hardware improvements, SR-MRI offers another possibility when the acquired data is not sufficient to characterize small but clinically significant structures. From the *k*-space perspective, CS is an interpolation problem whereas SR is an extrapolation one. Nonetheless, SR is actually considered as a post-acquisition processing, whose measurements are low resolution (LR) images in the spatial domain rather than undersampled *k*-space data. However, both CS-MRI and SR-MRI intend to learn an end-to-end nonlinear mapping from low quality observations to high quality reference images, and thus it is possible to extend ERRN framework into SR-MRI application. In optimization-based reconstructions, the image reconstruction can be solved using a mathematical model consisting of data fidelity and regularization terms, which may be in different domains. Here, we apply ERRN to SR-MRI reconstruction using back projection based error-correction instead of the data consistency for CS-MRI. Back projection is a well-known SR algorithm [47] that iteratively calculates the reconstruction error between the LR

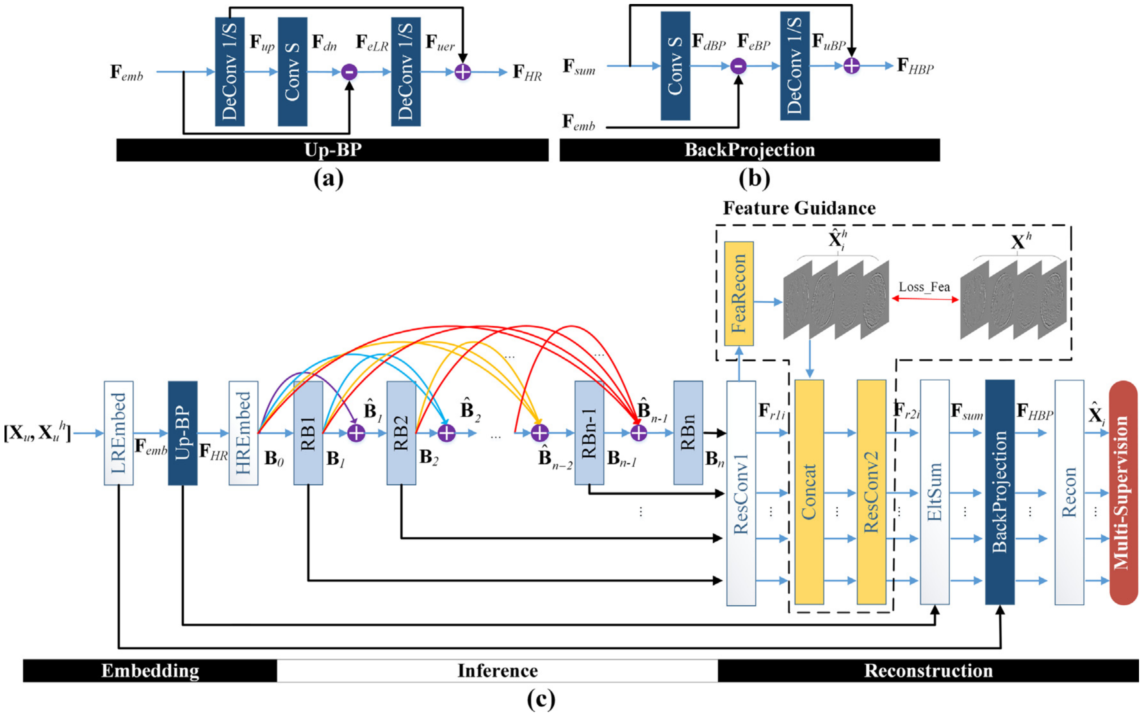

Fig. 4. ERRN network architecture for SR-MRI. (a) *Up-BP* block with the sampling rate $S \times S$; (b) *BackProjection* block; (c) ERRN framework. The components of inference net and feature guidance are same with those for CS-MRI in Fig. 3.

observation and its prediction, then projects the error back to refine the high resolution (HR) estimation. Inspired by the network in [36], we explore back projection in ERRN using mutually connected up- and down-sampling layers to learn the nonlinear relation between LR and HR images.

The ERRN framework modified for SR-MRI is presented in Fig. 4, which primary components are identical to that for CS-MRI in Fig. 3 except the error-correction module that is constituted by *Up-BP* block in embedding net and *BackProjection* block in reconstruction net. The back projection in *Up-BP* is executed using the following procedure: (1) upsampling the LR feature maps $\mathbf{F}_{emb}$ by deconvolutional layer *DeConv 1/S*; (2) downsampling the upscaled HR feature maps $\mathbf{F}_{up}$ through the convolutional layer *Conv S*, with sampling scale $S \times S$; (3) upsampling the residual error $\mathbf{F}_{eLR} = \mathbf{F}_{dn} - \mathbf{F}_{emb}$ in the LR feature space using another deconvolutional layer; (4) outputting the HR feature maps with $\mathbf{F}_{HR} = \mathbf{F}_{uer} + \mathbf{F}_{up}$. The elementwise summation $\mathbf{F}_{sum} = \mathbf{F}_{r2i} + \mathbf{F}_{HR}$ is input to *BackProjection* with the following procedure: (1) downsampling the HR feature maps $\mathbf{F}_{sum}$ with the convolutional layer and getting $\mathbf{F}_{dBP}$; (2) upsampling the residual error $\mathbf{F}_{eBP}$ in the LR feature space into $\mathbf{F}_{uBP}$ using the deconvolutional layer; (3) outputting the HR feature maps as $\mathbf{F}_{HBP} = \mathbf{F}_{uBP} + \mathbf{F}_{sum}$.

4. Experiment

4.1. MRI acquisition

ERRN performance was evaluated using a real-valued human brain dataset and a public complex-valued knee database [48]. To assess the overfitting problem of our ERRN network in generalization, e.g. in case of different sequence parameters, hardware imperfections, pathology, etc., pathological brain tumor data [49] and prospectively undersampled rat brain data were used to validate its generalization performance. The real-valued brain dataset contains 16 scans of 8 healthy adults acquired at different time using a 7T Philips Healthcare MRI equipped with a 32-channel head receive coil at the F. M. Kirby Research Center. The study was approved by the local IRB and written informed consent was obtained from the volunteers. For each scan, a T1-weighted sequence MPRAGE (magnetization prepared rapid gradient echo) was used with FOV = 220 mm × 220 mm × 110 mm, matrix size = 224 × 224 × 110, TR = 4 ms, TE = 2 ms, and flip angle = 7°. The MPRAGE sequence is characterized by excellent contrast between gray matter and white matter but, at high field, is hampered by B_1 inhomogeneity issues and, when using high resolution, by relatively low SNR and long scan time [50]. The public complex-valued knee dataset[1] contains 20 subjects acquired by fast spin echo sequence using a 3T GE MR750 whole body scanner. This dataset has been used in [27,28], and its detailed description is available in [48]. Some imaging parameters are FOV = 160 mm × 160 mm × 153 mm, matrix size = 320 × 320 × 256 and flip angle = 90°. The pathological data was obtained from BRATS 2018 (2018 Brain Tumor Image Segmentation Benchmark database)[2]. The BRATS dataset contains multi-institutional pre-operative multi-parametric MRI scans of 210 cases of glioblastoma, including T1-weighted, T1-weighted contrast-enhanced Gadolinium (T1Gd), T2-weighted and T2-weighted FLAIR clinical protocols. We used T1Gd images as testing data, with isotropic voxel size = 1 mm, matrix size = 240 × 240 × 155, and more acquisition information in [49]. The acquisition parameters for *in vivo* rat brain imaging is in our supplementary material.

4.2. Training

The real-valued brain data was preprocessed into magnitude images in the range of [0, 1]. We randomly selected 9 scans to generate the training data, while the remaining was grouped into 2 scans and 5 scans for validation and testing, respectively. For the real-valued brain dataset, 2D axial slices were used in training and testing. Following the experimental setup of knee data in

---

[1] http://old.mridata.org/fullysampled/knees.

[2] https://www.med.upenn.edu/sbia/brats2018/data.html.

[27], we selected out 10 sagittal slices in the middle of each case and then divided them into 2 sets, i.e. 100 slices for training and other 100 slices for testing. Each image was then rotated in 90° increments combined with a horizontal flip, to augment the dataset into 8 variants. The training data was not cropped into patches in CS-MRI application because the data consistency is operated in *k*-space. In SR-MRI, images were split into patches of size 68 × 68 comparable to the receptive field of ERRN.

In ERRN network, the number of residual blocks was optimized to be $n = 10$ in the experiment. The convolutional layers had a kernel size of 3 × 3 and the channel number of 64, and the kernel size of deconvolutional layers was defined as $(2S - mod(S,2)) \times (2S - mod(S,2))$ in back projection blocks, with S indicating the sampling scale $S \times S$. The convolutional kernel weights were initialized using the normal distribution $\mathcal{N}\left(0, \sqrt{2/kn^2 \times chn}\right)$, where kn denotes the kernel size and chn is the channel number. More detailed information about parameters setting is available in [46]. The learning rate was initialized to be 10^{-4} with minibatches of size 16. The network was trained separately for different undersampling schemes and rates with 100 epochs over 40 h per model on the Tensorflow [51] deep learning framework using ADAM optimizer [52]. The computing system is equipped with an Intel Xeon E5-2620 CPU, one 12 GB NVIDIA Pascal Titan X GPU, and 64 GB RAM.

4.3. Evaluation

In the first step, we evaluated the ERRN performance on CS-MRI reconstruction using undersampling patterns of radial [27] and variable-density random [42], at sampling rates of 10%, 20% and 30%, equal to acceleration factors of ×10, ×5 and × 3.3. The fully-sampled data was employed for the network label in training and the reference in result evaluation. Except for one case of prospectively undersampled rat brain data in our supplementary material, the undersampled data was generated from the fully-sampled data using various sampling schemes. Our ERRN was compared to conventional methods of sparsity-based TV [1] and non-local based PANO [34], and neural networks of U-net [23] and DC-CNN [21]. In addition, the zero-filling image was used to illustrate the undersampled image quality, artifacts severity and their structural characteristics. The conventional methods were implemented using their open source codes, while regularization parameters were set to be 10^6 and the predefined inner loop improvement tolerance was assigned as 0.05 following [34]. U-net and DC-CNN networks were all operated on the Tensorflow using the framework specifications in [23] and [21]. Reconstructed results were quantitatively assessed in terms of PSNR (peak signal-to-noise ratio) and MSSIM (mean structure similarity index map) [53]. With respect to the reference image, PSNR is calculated as $10\log_{10}(1/MSE)$, while MSSIM is a normalized parameter proportional to the structure similarity of two images. Higher PSNR and MSSIM scores mean better reconstruction.

Moreover, we demonstrated the ERRN performance on SR-MRI using the real-valued brain data. The undersampled images were synthesized by blurring fully-sampled data using Gaussian kernels of size 3 × 3, 5 × 5 and 7 × 7 with standard deviation $\sigma = 0.8$, 1.0, 1.6, and then downsampled at scales of 2 × 2, 3 × 3 and 4 × 4, respectively. ERRN was compared to conventional methods of LRTV based on low-rank and total variation [4] and ScSR based on sparse coding [5], and CNN-based methods of SRCNN [13] and VDSR [14]. The codes of LRTV and ScSR methods came from public resources, and their regularization parameters were determined to be 0.01 and 0.15, respectively. The dictionary size in ScSR was assigned as 1024 to maximize its presentation power. SRCNN and VDSR networks were implemented using their published codes in Caffe [54]. All networks were trained with the same training dataset and optimized by the ADAM optimizer for 100 epochs. According to a set of experiments at sampling rate 10%, the CS-MRI reconstruction time of ERRN (62.6 ms, GPU) was comparable to those of U-net (61.9 ms, GPU) and DC-CNN (36.4 ms, GPU). Meanwhile, the SR-MRI reconstruction time of ERRN (64.3 ms, GPU) was also comparable to those of VDSR (25.8 ms, GPU) at sampling scale 4 × 4. The reconstruction time was an average for all brain images in the testing data running with the same environment. By means of multi-supervision, ERRN executes the intermediate prediction for each residual block and then weighted averages them in the final reconstruction. Therefore, ERRN is somewhat time consuming compared to DC-CNN in spite of similar depths.

4.4. Network architecture

Table 1 lists architecture parameters of all networks used in our experiments. In CS-MRI reconstruction, the DC-CNN network cascades 5 subnetworks of each containing 5 layers, which depth is comparable to ERRN but the number of network parameters is more than 3 times that of ERRN. In contrast, the network of U-net has 23 convolutional layers and 31,024,384 parameters, far more than 165,952 in ERRN. As for the ERRN in SR-MRI, its depth is increased to 31 layers and the number of network parameters varies with the sampling scale $S \times S$ due to the usage of back projection blocks. In contrast, SRCNN contains only 3 convolutional layers and thus has fewest network parameters. VDSR equipped with 20 layers has 664,704 parameters, while ERRN has 356,288 at $S = 4$. Overall, ERRN has fewer network parameters than other networks of comparable depths.

5. Results

The ERRN framework was first analyzed to validate the optimal number of residual blocks and effects of three functional modules. Then, ERRN performance on undersampled reconstructions was mainly investigated in CS-MRI, followed with its extension in SR-MRI.

Table 1

Comparison in network architectures used in experiments (*S* denotes the scale factor).

Network		Depth	Kernel size	Channels	Parameters
CS-MRI	U-net	23	3 × 3/2 × 2//1 × 1	2/64/128/256/512/1024	31,024,384
	DC-CNN	25	3 × 3	2/64	564,480
	ERRN	27	3 × 3	64/8/2	165,952
SR-MRI	SRCNN	3	9 × 9/5 × 5	64/32/1	57,184
	VDSR	20	3 × 3	64/1	664,704
	ERRN	31	3 × 3	64/4/1	274,368 $S = 2$ 294,848 $S = 3$ 356,288 $S = 4$

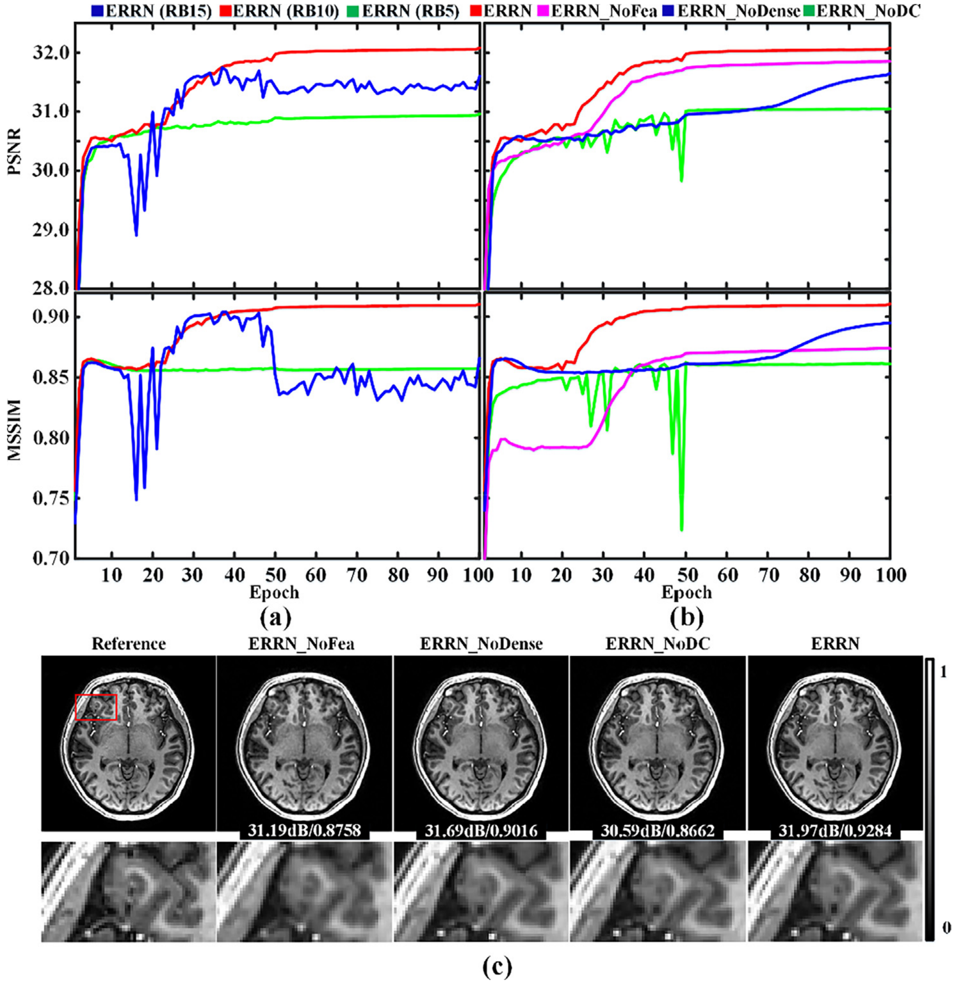

Fig. 5. ERRN network architecture analysis in CS-MRI reconstruction on brain validation data using 10% random sampling. The optimal number of residual blocks (a) and the effect of each functional module (b) are validated by PSNR-epoch and MSSIM-epoch curves. (c) Reconstruction results of modified ERRNs compared to the reference image.

5.1. Network architecture analysis

We analyzed ERRN framework using performance curves in terms of PSNR-epoch and MSSIM-epoch averaged in validation data. The optimal number of residual blocks is determined for ERRN in Fig. 5a, while the enhancements of feature guidance, dense connections and data consistency are verified in Fig. 5b and c. Since an aggressive undersampling rate over 5-fold is usually challenging, Fig. 5 illustrates the performance curves of 10% random sampling on real-valued brain CS-MRI reconstruction. In general, the effects of functional modules at different sampling patterns and sampling rates are consistent in undersampled MRI reconstructions. Corresponding evaluations of SR-MRI are given in Fig. S3.

5.1.1. Number of residual blocks

To validate how ERRN can benefit from the increased depth, we tested the number of residual blocks in ERRN network, denoted as ERRN (RB5/10/15) for $n = 5$, 10 and 15. In Fig. 5a, performance curves gradually converge after 50 epochs, and ERRN (RB10) indeed boosts the reconstruction compared to ERRN (RB5), whereas ERRN (RB15) performs worse than ERRN (RB10). Therefore, we determined to set $n = 10$ for ERRN in experiments. The test result in ERRN indicates that recursive learning may limit the optimal depth of residual networks. Although as the network becomes deeper, it can cover a larger receptive field that could obtain more information and then improve the network performance. Nevertheless, we think the sensitivity of network performance to the number of blocks depends on the specific network architecture.

5.1.2. Effects of functional modules

We also analyzed the effects of three functional modules by comparing ERRN with its corresponding modifications of ERRN_NoDense, ERRN_NoFea and ERRN_NoDC, in which a single module was removed from ERRN network for a clear evaluation. Fig. 5b illustrates their PSNR-epoch and MSSIM-epoch curves, while Fig. 5c shows reconstructed images. Three functional modules are all proved to have noticeable impacts on CS-MRI reconstruction results. Among them, the improvement of data consistency is the most significant, for ERRN_NoDC exhibits 1 dB decrease in PSNR and 0.05 in MSSIM. Meanwhile, fine structures in the ERRN_NoFea result still lack clarity. Interestingly, we can observe that the feature guidance produces a more obvious increase in MSSIM than the increase of PSNR, which agrees with our purpose that feature guidance is designed to learn high-frequency features so that ERRN can achieve sharper contrast. Lastly, the perfomance curves of ERRN_NoDense cannot converge within 100 epochs, which means that dense connections could benefit ERRN with a stable convergence.

5.2. CS-MRI reconstruction

In this section, results on real-valued brain data and complex-valued knee data were generated from the network trained by the corresponding dataset. In order to assess the generalization capability of ERRN network, we reconstructed pathological brain images by directly employing the network parameters trained from the healthy real-valued brain experiment without transfer learning, which can be called as a zero-shot inference problem [26]. In addition, we also acquired prospectively accelerated data of *in vivo* rat brain and performed its reconstruction experiment using the ERRN network trained from the knee data in axial slices. The reconstruction results of prospectively undersampled data are presented in Fig. S9 in our supplementary material.

5.2.1. Results on real-valued brain data

Figs. 6 and 7 show real-valued brain testing results in a representative axial slice using radial and random samplings at rates 10% and 20%, associated with image PSNR and MSSIM values, zooming views and absolute difference maps. The reconstructions at sampling rate 30% are provided in Figs. S4 and S5, and we also illustrate the results at rate 10% in coronal and sagittal views in Fig. S6. All results demonstrate the excellent performance of ERRN across varying sampling schemes.

In Fig. 6, the zero-filling image of radial sampling at rate 10% represents severe blurring and streaking artifacts, whereas moderate at rate 20%. For the sampling rate 10%, the conventional method TV presents obvious washing effects and fails to obtain a reasonable reconstruction. The PANO result appears slightly better but suffers from noticeable edges blurring and structures missing. In contrast, results of CNN-based methods are improved qualitatively and quantitatively. However, the U-net reconstruction remains unclear and noisy, inferior to other two networks. In the DC-CNN result, tissue structures are recovered better than the U-net but some fine features are lost, as indicated by the arrow in zooming views. ERRN reconstruction can recover the finest details and the sharpest contrast, with the least undersampling artifact and the highest PSNR/MSSIM scores. Besides, there are smallest errors in the difference map of ERRN. As for sampling rate 20%, the performances of all methods are coherent with those of rate 10%, but the reconstructed images have higher quality.

In Fig. 7, each method show better reconstructions in random sampling than the corresponding results of radial sampling in Fig. 6, with accurate tissue structures and increased quantitative scores. In particular, ERRN reconstructions are perceptually identical to the reference image, with no obvious error in difference maps. Table 2 lists the averaged PSNR and MSSIM of all testing images at various sampling rates and sampling masks. ERRN always achieves the best performance and has the highest scores of the image quality metrics. For instance, the ERRN result at sampling rate 10% has increased PSNR by 3 dB than conventional method PANO, while about 0.2 in MSSIM. Even for state-of-the-art networks, ERRN outperforms DC-CNN and U-net at different sampling rates with higher PSNR and MSSIM scores.

5.2.2. Results on complex-valued knee data

Fig. 8 illustrates reconstructed magnitude and phase images of knee data using 10% radial sampling, associated with zooming views and absolute difference maps of magnitude images. Similar to the real-valued brain experiments, radial sampling results in serious streaking artifacts, especially for the phase image. Conventional methods of TV and PANO are not able to gain convincing reconstructions at such aggressive undersampling rate, whereas U-net even performs worse than PANO. However, ERRN and DC-CNN methods can suppress most artifacts in magnitude and phase images, but DC-CNN results are still a bit blurred at edges and

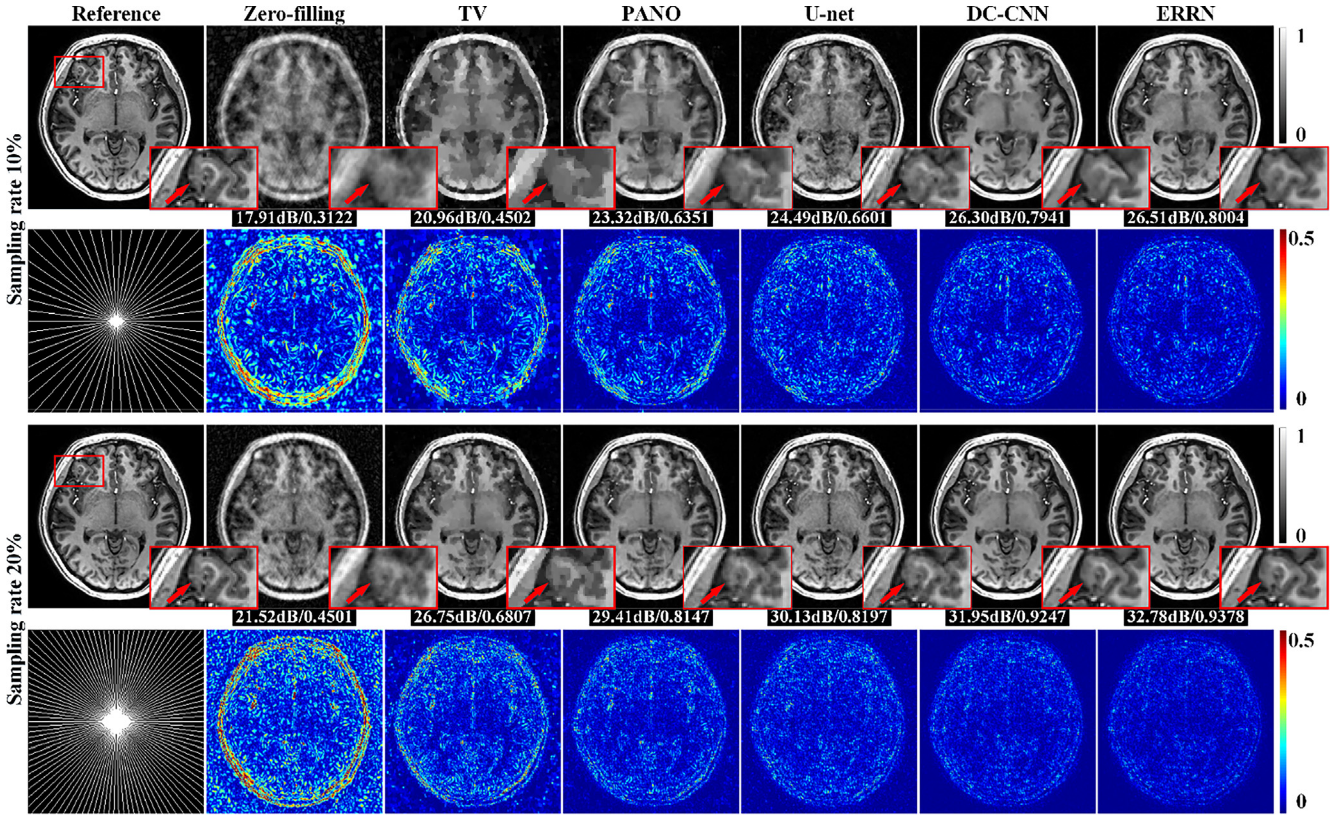

Fig. 6. CS-MRI reconstructions on real-valued brain images using radial sampling at rates 10% and 20%. The first column shows the fully-sampled reference image and sampling masks. The second to last columns are results of different methods, with zooming views and absolute difference maps. The red arrow indicates the region of remarkable distinctions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

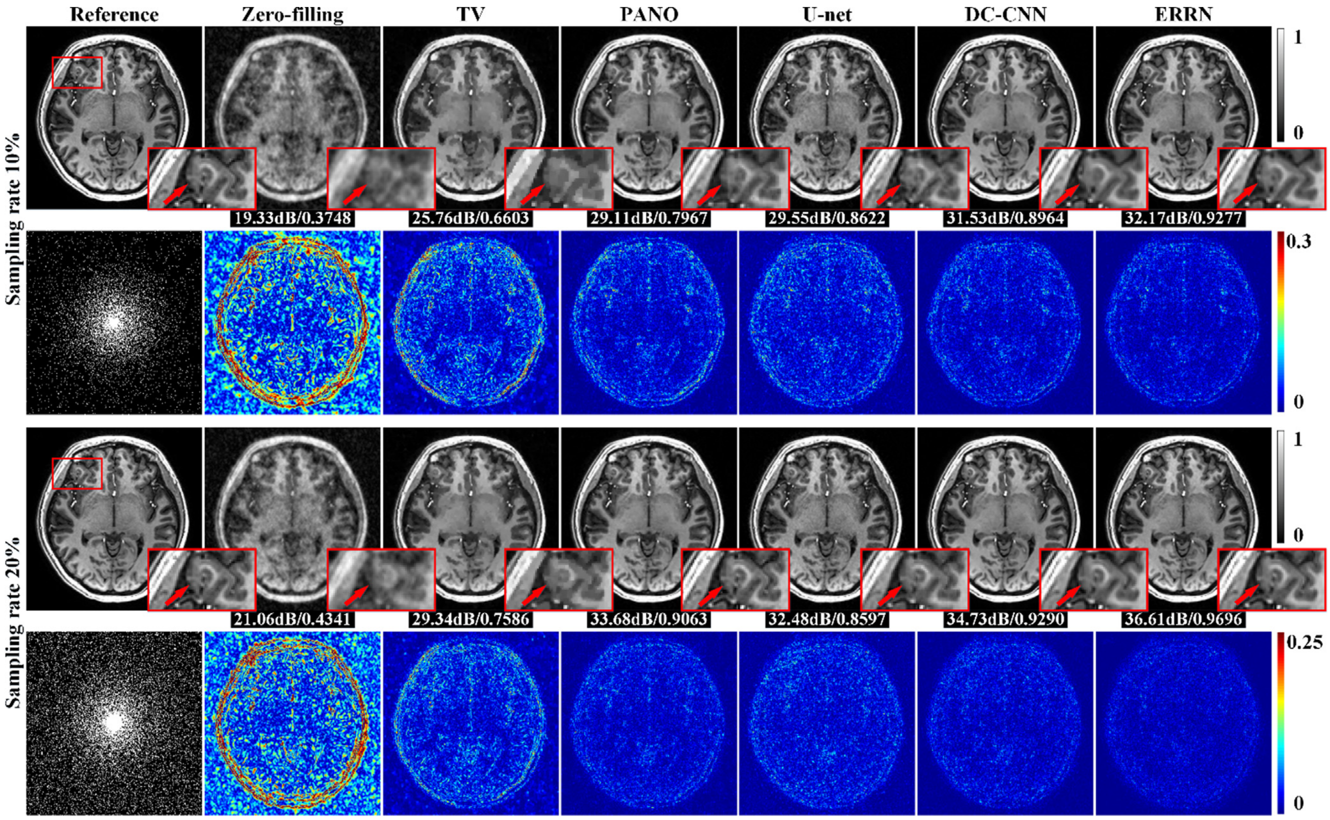

Fig. 7. CS-MRI reconstructions on real-valued brain images using random sampling at rates 10% and 20%. The first column shows the fully-sampled reference image and sampling masks. The second to last columns are results of different methods, with zooming views and absolute difference maps. The red arrow indicates the region of remarkable distinctions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Quantitative comparison of CS-MRI reconstructions on real-valued brain data in terms of PSNR (dB, top raw) and MSSIM (bottom raw).

Method	10%		20%		30%	
	Radial	Random	Radial	Random	Radial	Random
Zero-filling	18.51 ± 0.51 0.3042 ± 0.0154	19.79 ± 0.55 0.3665 ± 0.0169	21.83 ± 0.52 0.4350 ± 0.0182	21.44 ± 0.51 0.4211 ± 0.0181	24.39 ± 0.62 0.5127 ± 0.0203	23.31 ± 0.53 0.4800 ± 0.0195
TV	21.27 ± 0.64 0.4423 ± 0.0195	26.07 ± 0.9752 0.6515 ± 0.0292	26.72 ± 1.11 0.6635 ± 0.0304	29.32 ± 1.14 0.7335 ± 0.0261	31.32 ± 1.22 0.8236 ± 0.0292	33.13 ± 1.28 0.8287 ± 0.0249
PANO	23.61 ± 1.08 0.6222 ± 0.0385	29.46 ± 1.28 0.7857 ± 0.0268	29.69 ± 1.34 0.8196 ± 0.0340	34.10 ± 1.34 0.9055 ± 0.0196	33.72 ± 1.28 0.9022 ± 0.0208	38.16 ± 1.32 0.9492 ± 0.0116
U-net	24.71 ± 1.39 0.6454 ± 0.0434	30.08 ± 1.25 0.8621 ± 0.0270	30.17 ± 1.44 0.8067 ± 0.0292	32.89 ± 1.32 0.8537 ± 0.0214	33.42 ± 1.32 0.8769 ± 0.0211	36.37 ± 1.35 0.9068 ± 0.0162
DC-CNN	26.56 ± 1.45 0.7888 ± 0.0345	31.87 ± 1.40 0.8847 ± 0.0164	32.17 ± 1.44 0.9240 ± 0.0116	35.02 ± 1.35 0.9269 ± 0.0101	36.92 ± 1.34 0.9663 ± 0.0046	39.72 ± 1.31 0.9614 ± 0.0066
ERRN	**26.88 ± 1.58** **0.8016 ± 0.0325**	**32.28 ± 1.45** **0.9286 ± 0.0133**	**32.95 ± 1.54** **0.9397 ± 0.0101**	**36.90 ± 1.50** **0.9708 ± 0.0061**	**37.31 ± 1.37** **0.9759 ± 0.0040**	**41.45 ± 1.45** **0.9876 ± 0.0028**

somewhat over-smoothed as shown in zooming views. In addition, errors in the difference map of DC-CNN are more obvious over tissue boundaries. ERRN also can obtain outstanding performances in other cases, e.g., results of 10% random sampling in Fig. S7 where ERRN magnitude result presents an excellent similarity with the reference image.

Table 3 lists the quantitative comparison on the magnitude and phase results of two sampling patterns at rate 10%. All evaluation values are the average of all testing images. These quantitative evaluations are consistent with the figure results, and obviously ERRN achieves the best PSNR and MSSIM scores. As demonstrated in real-valued brain data, the reconstructions here on random sampling surpass radial sampling for the magnitude image, which is in agreement with the results in real-valued brain experiment. On the other side, magnitude images have higher quality than phase images in both the zero-filling and other reconstruction results. That is because the phase data is naturally more sensitive to imaging perturbations and noise. Moreover, the real and imaginary components of complex-valued data are processed in separate network channels in our experiment, and any small bias in the imaginary part may be amplified in the phase calculation.

5.2.3. Results on pathological brain tumor data

In Fig. 9, the results of pathological data are reconstructed using the ERRN network trained on healthy real-valued brain data, and no pathological image was used for network training. We ran-

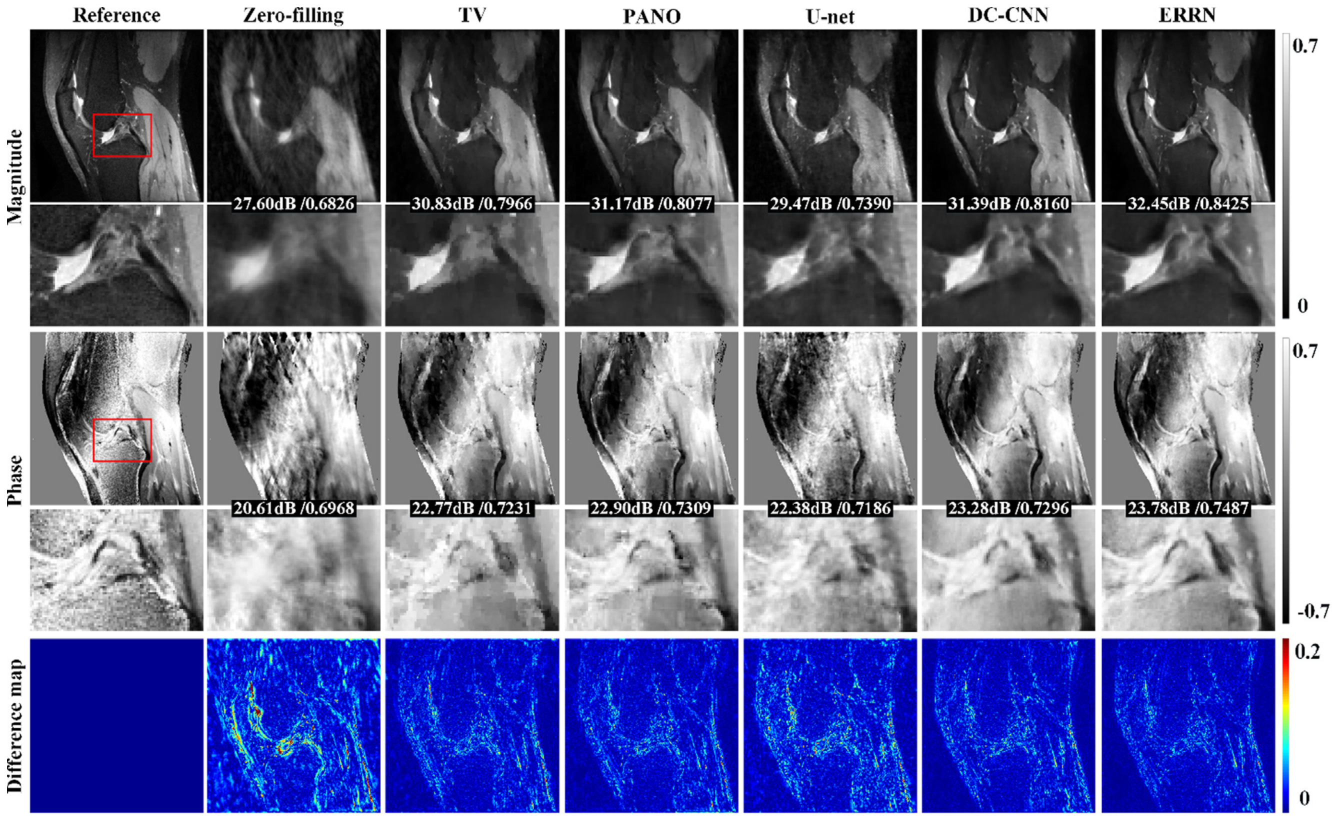

Fig. 8. CS-MRI reconstructions on complex-valued knee data using radial sampling at rate 10%. The first column shows reference images, and the second to last columns are results of different methods. From the top to bottom are magnitude and phase images with zooming views and absolute difference maps.

Table 3
Quantitative comparison of CS-MRI reconstructions on complex-valued knee data at sampling rate 10% in terms of PSNR (dB, top raw) and MSSIM (bottom raw).

Method	Radial sampling		Random sampling	
	Magnitude	Phase	Magnitude	Phase
Zero-filling	27.48 ± 0.97 0.6619 ± 0.0367	20.15 ± 1.31 0.6695 ± 0.0513	28.69 ± 0.98 0.7034 ± 0.0356	18.79 ± 1.17 0.6463 ± 0.0553
TV	30.87 ± 1.28 0.7850 ± 0.0308	22.28 ± 1.39 0.7052 ± 0.0494	33.09 ± 1.37 0.8377 ± 0.0265	22.40 ± 1.31 0.7162 ± 0.0504
PANO	31.27 ± 1.34 0.7998 ± 0.0307	22.63 ± 1.37 0.7191 ± 0.0464	33.70 ± 1.41 0.8542 ± 0.0241	22.95 ± 1.29 0.7347 ± 0.0463
U-net	29.21 ± 1.11 0.7142 ± 0.0351	21.68 ± 1.33 0.6891 ± 0.0506	31.68 ± 1.21 0.7910 ± 0.0300	21.46 ± 1.30 0.6966 ± 0.0510
DC-CNN	31.45 ± 1.34 0.8097 ± 0.0275	22.73 ± 1.30 0.7112 ± 0.0488	34.08 ± 1.48 0.8665 ± 0.0223	23.26 ± 1.32 0.7313 ± 0.0470
ERRN	**31.96 ± 1.49** **0.8245 ± 0.0290**	**23.23 ± 1.37** **0.7306 ± 0.0460**	**34.57 ± 1.59** **0.8788 ± 0.0232**	**23.61 ± 1.36** **0.7472 ± 0.0450**

domly selected 15 cases with glioblastoma from the database, and in each case tested on 40 T1Gd images with lesion information in axial slices. Fig. 9 shows ERRN reconstructions on the pathological data accompanied with absolute difference maps and zooming views of the tumor region. For radial and random samplings at rates 20% and 30%, ERRN achieves faithful reconstructions and retains clear pathological patterns, i.e. the morphology of brain tumors is exactly restored without any distortion or any synthetization. As for the rate 10%, ERRN performs well in random sampling, while the result of radial sampling is acceptable but could be further improved, which is in agreement with the previous demonstration that CS-MRI reconstruction on 10% radial sampling is more difficult. Table S1 in our supplementary material lists PSNR and MSSIM scores averaged in all pathological reconstructions on different sampling rates and sampling patterns. This experiment indicates that our proposed ERRN can reduce the overfitting problem in network generalization.

5.3. SR-MRI reconstruction

Fig. 10 displays SR-MRI reconstructions at downsampling scales 3 × 3 and 4 × 4 on real-valued brain images, and results at scale 2 × 2 are arranged as Fig. S8 in supplementary material. We can see that the CNN-based methods have remarkable improvements over the conventional methods, yielding higher image contrast, SNR and resolution, simultaneously. Compared to the simple

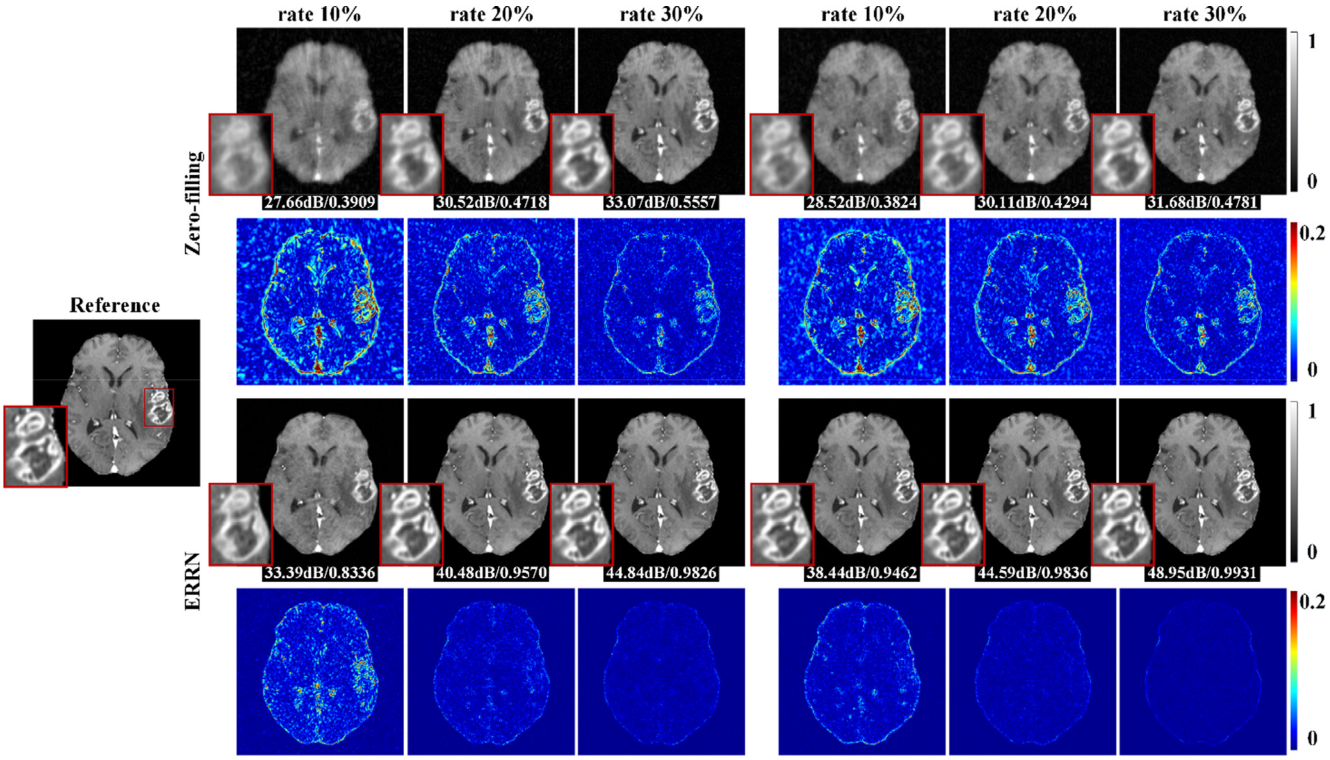

Fig. 9. CS-MRI reconstructions on the pathological brain image. The reference image is in the left. First three columns are results of radial sampling while right three columns are random sampling results. From top to bottom are zero-filling images and ERRN results, accompanied with zooming views and absolute difference maps.

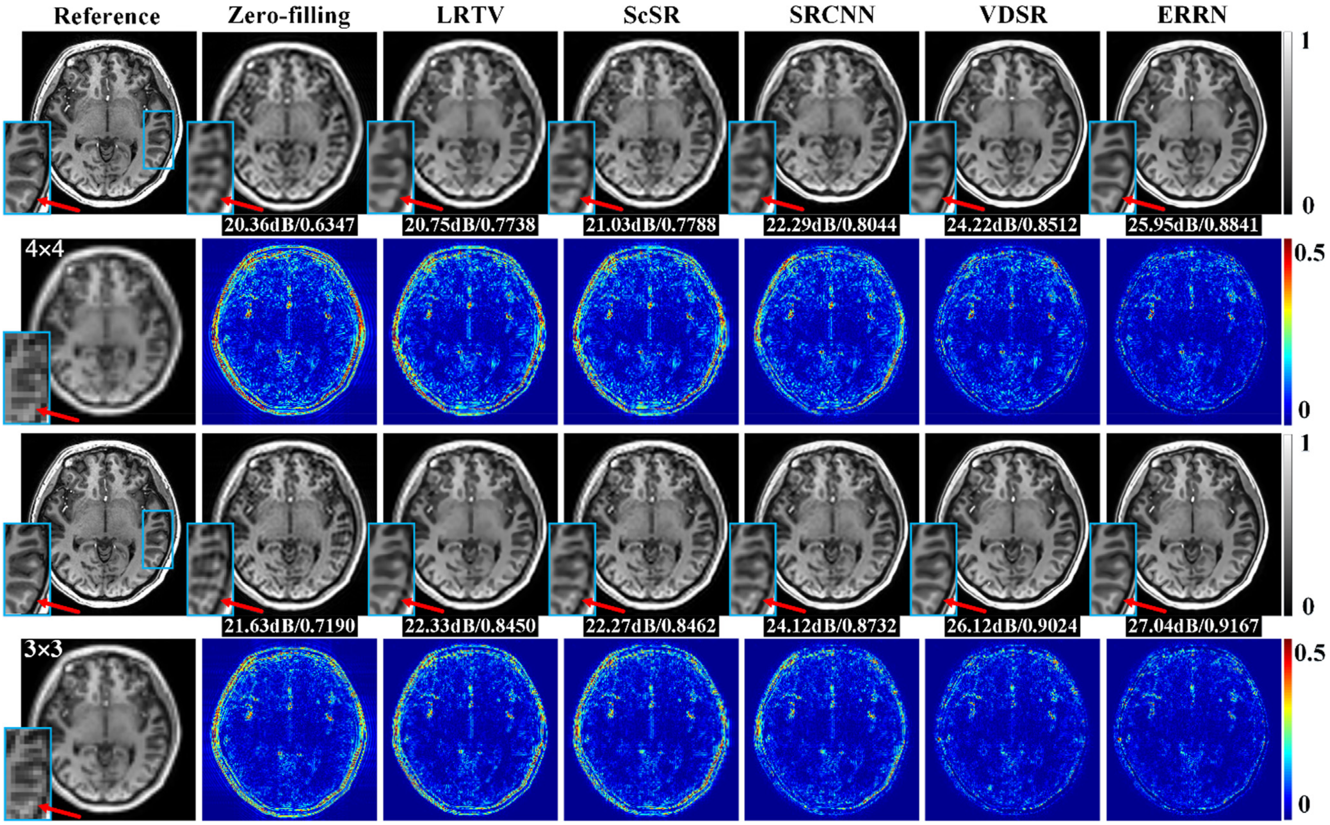

Fig. 10. SR-MRI reconstructions on real-valued brain data at scales 4 × 4 and 3 × 3 with zooming views and absolute difference maps. The first column contains the reference image and undersampled low resolution images. From the second to last columns are the results of different methods.

Table 4
Quantitative comparison of SR-MRI reconstructions on real-valued brain data in terms of PSNR (dB, top raw) and MSSIM (bottom raw).

Method	2 × 2	3 × 3	4 × 4
Zero-filling	25.50 ± 1.06 0.8245 ± 0.0297	21.87 ± 0.97 0.6992 ± 0.0342	20.05 ± 0.91 0.6095 ± 0.0401
LRTV	25.19 ± 1.03 0.9141 ± 0.0180	22.25 ± 0.95 0.8395 ± 0.0343	20.69 ± 0.89 0.7709 ± 0.0414
ScSR	26.07 ± 1.09 0.9343 ± 0.0137	22.38 ± 0.98 0.8444 ± 0.0323	20.97 ± 0.90 0.7736 ± 0.0428
SRCNN	27.99 ± 1.21 0.9445 ± 0.0128	23.99 ± 1.18 0.8693 ± 0.0313	22.07 ± 1.01 0.7990 ± 0.0438
VDSR	29.53 ± 1.20 0.9556 ± 0.0108	25.97 ± 1.36 0.9012 ± 0.0267	24.09 ± 1.38 0.8501 ± 0.0398
ERRN	**31.18 ± 1.23** **0.9657 ± 0.0091**	**27.04 ± 1.43** **0.9192 ± 0.0239**	**26.12 ± 1.29** **0.8904 ± 0.0308**

SRCNN, the deep networks of VDSR and ERRN achieve better performances, but our proposed ERRN obtains the best PSNR and MSSIM values. Especially as the downsampling scale growing, the low resolution image becomes so severely corrupted that the conventional methods and SRCNN fail to restore an effective reconstruction. In contrast, ERRN is always able to generate an excellent restoration both perceptually and quantitatively. In particular, tissue structures in the ERRN results are closest to the reference, e.g. the boundaries between gray matter, white matter and cerebrospinal fluid, the contours of cerebral gyri and blood vessels. The cortex gyrus noted by the red arrow is well recovered in our network results, whereas remains blurred in other methods. Table 4 lists the mean values of PSNR and MSSIM in real-valued brain testing data. The PSNR at scale 4 × 4 increases from 20.97 dB in ScSR to 26.12 dB in ERRN, while MSSIM rises from 0.77 to 0.89. Compared to VDSR, ERRN gains an increase of 1 dB-2 dB in PSNR and 0.01–0.04 in MSSIM at scales of 2 × 2 to 4 × 4. Although VDSR also exploits the residual learning in a feed forward architecture, ERRN surpasses it by a noticeable margin due to the error-correction and feature guidance modules.

6. Discussion

Theoretical works have already justified the relationship between convolutional networks and conventional method [13,17,33,55], which provide solid support for CNN-based method development. In this work, we focus on how to build a powerful network that enables outstanding performance in undersampled MRI reconstructions especially for aggressive undersampling rates. Thus, a deep network named as ERRN is proposed and demonstrated in CS-MRI and SR-MRI applications with various data and datasets. Constituted by three subnets, ERRN framework has a flexible architecture to incorporate functional modules. The network input is firstly preprocessed in the embedding net, and then signals flow to the inference net which learns residual information with a group of recursive residual blocks. The theoretical basis of recursive residual network for image reconstruction is elaborated in supplementary material. Attributing to recursive learning, parameter-shared residual blocks can boost network performance through enlarging the receptive field without adding new parameters. Besides, dense connections are introduced to the inference net to strengthen information flow. Compared to concatenating feature maps at each node, the scalar weights learned in network training are used to improve the dense connections in ERRN to prevent parameters explosion.

To fully exploit the image morphology, we developed a high-frequency feature guidance based on a priori extracted from label data with four directional filters. The feature guidance module is added into the reconstruction net and utilized to predict the underlying anatomy, playing a complementary role to the residual learning. As demonstrated in Figs. 2 and 5, ERRN benefits from this functional module with improved structure sharpness, even if the undersampled image severely suffers from various artifacts and blurring. For example, DC-CNN also employs residual blocks and data consistency in the network, but compared to DC-CNN, ERRN has significant improvements on MSSIM and image visualization thanks to the feature guidance. Here, the directional filters were created to detect structural features along different orientations, whereas there could be other options according to image characteristics. Overall, the feature guidance can effectively constrain the reconstructed image in the solution space and strive to reflect accurate tissue structures.

Inspired by data fidelity term used in optimization-based methods, we enhance our network with an application-specific error-correction unit, which is inserted into the embedding and reconstruction nets. The error-correction unit can correct the intermediate prediction of each residual block RB$_i$ by retaining the useful information in undersampled data, and its effect has been validated to be significant for ERRN performance. In CS-MRI application, we implemented the data consistency operation as expressed in Eq. (5), which executes a linear fitting between the network predictions and original measurements in k-space, with the weighting parameter λ learned from training. Supposing an infinite weight for λ, the reconstruction result is equal to the original measurement for the acquired index, while the unacquired one is determined by ERRN network. Combined with dense connections and error-correction unit, the prediction of each recursive residual block in ERRN is somewhat analogous to the conventional iteration in optimization-based methods, but is more advanced and powerful. More insights can be seen in [20].

CS-MRI and SR-MRI are ill-posed inverse problems which need to reconstruct high quality images from low quality measurements. CNN-based approaches can solve these inverse problems by learning an end-to-end mapping from low quality observations to high quality label dataset [17,42]. For optimization-based reconstructions, the data fidelity term in mathematical models may have different forms according to the specific sampling strategy, whereas regularization constraints are often based on similar techniques, e.g. total variation, sparse a priori, neighbor embedding, etc. Hence, if we make some adjustments to the data fidelity module, ERRN framework is able to be adapted for both CS-MRI and SR-MRI applications. In this case, an adaptive error-correction unit is created to introduce physically meaningful data fidelity operations, i.e. data consistency for CS-MRI and back projection for SR-MRI as shown in Figs. 3 and 4, respectively. Meanwhile, the network input also depends on the specific application. Overall, a flexible architecture enables the network a good potential for application extension.

Considering that the raw data acquired in MRI is normally complex-valued whereas real-valued magnitude images are widely used in research and clinical applications, ERRN was evaluated on real- valued brain data and complex-valued knee data, i.e. two organs with different structural complexity. Our experiments focused on the CS-MRI reconstruction at sample rates 10% and 20%, because those are hard to deal with using optimization-based methods. In views of various undersampling schemes, we used radial and random patterns in our CS-MRI experiments, for they are appropriate to conduct aggressive undersampling rates. In addition, CS-MRI results on sampling rate 30% and most of SR-MRI related contents are provided in supplementary material. For complex-valued data, ERRN network processes the real and imaginary components in two separate channels both using real-valued convolutional kernels. However, the work in [56] developed a complex fully convolutional neural network for MR image reconstruction by introducing complex convolution and demonstrated its improved perceptual quality and recovery of anatomical structures in contrast to its real-valued counterparts. We think the complex convolutional kernels would be helpful when the data to be solved is complex-valued, and thus we would like to have some attempts on it in future.

Normally, when training data are not sufficient or lack of fully-sampled reference for the task, the supervised learning paradigm may break down to train a reliable model. Moreover, the model trained using a certain dataset often lacks the ability to generalize to other conditions. Although transfer learning can leverage the existing network parameters and then fine tunes them using a small specific dataset, improving network generalization capability offers another solution. To assess the ERRN performance in generalization, we applied the network trained by healthy real-valued brain dataset to reconstruct pathological brain images directly, without any transfer learning. Furthermore, we also acquired one case of prospectively Cartesian undersampled imaging of *in vivo* rat brain and then implemented its reconstruction by ERRN network trained from human knee dataset, as shown in Fig. S9. These two experimental results demonstrated that our ERRN network can effectively reduce the overfitting problem and generalize well because of the functional modules and fewer parameters.

In this work, we have proved that more control on the "black box" operation of deep learning is meaningful to design an effective application-specific network to avoid overfitting and achieve superior performance. However, there are usually certain balances between the network depth and its performance, as well as network overfitting and its parameters number, in framework development. If we arrange more residual blocks in ERRN without recursive learning, the reconstruction may be better, whereas the number of network parameters will increase and the generalization capability will decrease. When determining the optimal number of residual blocks, we can see in Fig. 5 that ERRN (RB10) indeed boosts the reconstruction compared to ERRN (RB5), whereas ERRN (RB15) performs worse than ERRN (RB10). It reveals that the current usage of recursive residual blocks has limit in training a deeper network. To address this issue, we have explored to arrange recursive residual blocks into the cascade style as [12] and [21], but that might lead to the sacrifice of generalization and beyond the scope of this paper.

Our experimental results show that the proposed ERRN can achieve promising results for 2D image reconstruction using 2D convolutional kernels. However, some MRI data acquired using 3D imaging sequence have to be processed with 3D convolutional network, e.g. quantitative susceptibility mapping. Although translating a 2D network into a 3D model has challenges in computational complexity, memory allocation and convergence stability, etc., it has already been well developed and widely applied [31,57,58]. Considering that parallel imaging becomes a standard feature on most MRI scanners, high quality reconstruction of undersampled multi-channel MR data is one of research interests in MRI deep learning. In the first step, coil sensitivity maps are required to be precomputed. The measured raw data of each channel, coil sensitivity maps, and zero-filling initializations should all be fed into the network as in [20]. Therefore, the realistic size of network input becomes much larger inevitably. In this case, an easy way is to add a concatenation layer in the embedding net of ERRN, while the feature guidance module and error-correction unit are updated with the coil-sensitivity weighted data.

7. Conclusion

Motivated by the outstanding performance of deep learning, we propose a network ERRN for undersampled MR image reconstruction. The ERRN is based on a recursive residual network and enhanced by a set of user-designed functional modules, i.e. feature guidance, error-correction unit and improved dense connections. Experimental results demonstrated that ERRN can achieve excellent performances in both CS-MRI and SR-MRI applications, with highest PSNR and MSSIM scores and distinctly restored anatomical structures compared to conventional optimization-based methods and other convolutional neural networks. Using both retrospectively and prospectively undersampled data with various sampling schemes, we also demonstrated the good performance of ERRN on reducing the overfitting problem in generalization, e.g. preservation of unique pathologies that are not included in the training dataset.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China under Grant 81301277, U1632274, Natural Science Foundation of Fujian Province of China 2019J01047 and National Institutes of Health NIH/NIBIB grant of USA P41EB015909.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmr.2019.07.020.

References

[1] M. Lustig, D. Donoho, J. Pauly, Sparse MRI: the application of compressed sensing for rapid MR imaging, Magn. Reson. Med. 58 (2007) 1182–1195.
[2] E. Plenge, D.H. Poot, M. Bernsen, G. Kotek, G. Houston, P. Wielopolski, D.W.L. Van, W.J. Niessen, E. Meijering, Super-resolution methods in MRI: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?, Magn Reson. Med. 68 (2012) 1983–1993.
[3] A. Gholipour, J.A. Estroff, S.K. Warfield, Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI, IEEE Trans. Med. Imag. 29 (2010) 1739–1758.
[4] S. Feng, C. Jian, W. Li, P.T. Yap, D. Shen, LRTV: MR image super-resolution with low-rank and total variation regularizations, IEEE Trans. Med. Imag. 34 (2015) 2459–2466.
[5] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (2010) 2861–2873.
[6] S. Ravishankar, Y. Bresler, MR image reconstruction from highly undersampled k-space data by dictionary learning, IEEE Trans. Med. Imag. 30 (2011) 1028–1041.
[7] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Proc. NIPS (2012) 1097–1105.
[8] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE CVPR (2016) 770–778.
[9] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, Proc. IEEE CVPR (2017) 2261–2269.
[10] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, Proc. MICCA I (2015) 234–241.
[11] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep CNN denoiser prior for image restoration, Proc. IEEE CVPR (2017) 2808–2817.
[12] Y. Tai, J. Yang, X. Liu, C. Xu, MemNet: a persistent memory network for image restoration, Proc. IEEE ICCV (2017) 4549–4557.
[13] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 295–307.
[14] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, Proc. IEEE CVPR (2016) 1646–1654.
[15] J. Kim, J. Kwon Lee, K. Mu Lee, Deeply-recursive convolutional network for image super-resolution, Proc. IEEE CVPR (2016) 1637–1645.
[16] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, Proc. IEEE CVPR (2017) 105–114.
[17] K.H. Jin, M.T. Mccann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, IEEE Trans. Image Process. 26 (2017) 4509–4522.
[18] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, D. Liang, Accelerating magnetic resonance imaging via deep learning, Proc. IEEE ISB I (2016) 514–517.
[19] Y. Yang, J. Sun, H. Li, Z. Xu, Deep ADMM-Net for compressive sensing MRI, Proc. NIPS (2016) 10–18.
[20] K. Hammernik, T. Klatzer, E. Kobler, M.P. Recht, D.K. Sodickson, T. Pock, F. Knoll, Learning a variational network for reconstruction of accelerated MRI data, Magn. Reson. Med. 79 (2017) 3055–3071.

[21] J. Schlemper, J. Caballero, J.V. Hajnal, A. Price, D. Rueckert, A deep cascade of convolutional neural networks for dynamic MR image reconstruction, IEEE Trans. Med. Imag. 37 (2018) 491–503.

[22] D. Lee, J. Yoo, S. Tak, J.C. Ye, Deep residual learning for accelerated MRI using magnitude and phase networks, IEEE Trans. Biomed. Eng. 65 (2018) 1985–1995.

[23] D. Lee, J. Yoo, J.C. Ye, Deep residual learning for compressed sensing MRI, Proc. IEEE ISB I (2017) 15–18.

[24] C.M. Hyun, H.P. Kim, S.M. Lee, S. Lee, J.K. Seo, Deep learning for undersampled MRI reconstruction, Phys. Med. Biol. 63 (2018) 135007.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Proc. NIPS (2014) 2672–2680.

[26] G. Yang, S. Yu, H. Dong, G. Slabaugh, P.L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, D. Firmin, DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction, IEEE Trans. Med. Imag. 37 (2018) 1310–1321.

[27] T.M. Quan, T. Nguyen-Duc, W. Jeong, Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss, IEEE Trans. Med. Imag. 37 (2018) 1488–1497.

[28] M. Mardani, E. Gong, J.Y. Cheng, S.S. Vasanawala, G. Zaharchuk, L. Xing, J.M. Pauly, Deep generative adversarial neural networks for compressive sensing MRI, IEEE Trans. Med. Imag. 38 (2019) 167–179.

[29] O. Oktay, W. Bai, M. Lee, R. Guerrero, K. Kamnitsas, J. Caballero, A. de Marvao, S. Cook, D.O. Regan, D. Rueckert, Multi-input cardiac image super-resolution using convolutional neural networks, Proc. MICCA I (2016) 246–254.

[30] J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, H. Xu, Super-resolution reconstruction of MR image with a novel residual learning network algorithm, Phys. Med. Biol. 63 (2018) 085011.

[31] A.S. Chaudhari, Z. Fang, F. Kogan, J. Wood, K.J. Stevens, E.K. Gibbons, J.H. Lee, G. E. Gold, B.A. Hargreaves, Super-resolution musculoskeletal MRI using deep learning, Magn. Reson. Med. 80 (2018) 2139–2154.

[32] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S.A. Cook, A.D. Marvao, T. Dawes, R.B. Kainz, B. Glocker, D. Rueckert, Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation, IEEE Trans. Med. Imag. 37 (2018) 384–395.

[33] B. Zhu, J.Z. Liu, S.F. Cauley, B.R. Rosen, M.S. Rosen, Image reconstruction by domain-transform manifold learning, Nature 555 (2018) 487–492.

[34] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, Z. Chen, Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator, Med. Image Anal. 18 (2014) 843–856.

[35] E.M. Eksioglu, Decoupled algorithm for MRI reconstruction using nonlocal block matching model: BM3D-MRI, J Math. Imaging Vis. 56 (2016) 430–440.

[36] M. Haris, G. Shakhnarovich, N. Ukita, Deep back-projection networks for super-resolution, Proc. IEEE CVPR (2018) 1664–1673.

[37] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, S. Yan, Deep edge guided recurrent residual learning for image super-resolution, IEEE Trans. Image Process. 26 (2017) 5895–5907.

[38] A. Singh, N. Ahuja, Sub-band energy constraints for self-similarity based super-resolution, Proc. IEEE ICPR (2014) 4447–4452.

[39] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, Proc. IEEE CVPR (2017) 2790–2798.

[40] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv:1502.03167 (2015) https://arxiv.org/abs/1502.03167.

[41] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, Proc. ICML (2010) 807–814.

[42] H.K. Aggarwal, M.P. Mani, M. Jacob, MoDL: model-based deep learning architecture for inverse problems, IEEE Trans. Med. Imag. 38 (2019) 394–405.

[43] T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, Proc. IEEE ICCV (2017) 4799–4807.

[44] D.S. Weller, S. Ramani, J.A. Fessler, Augmented Lagrangian with variable splitting for faster non-Cartesian L1-SPIRiT MR image reconstruction, IEEE Trans. Med. Imag. 33 (2014) 351–361.

[45] H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, IEEE Trans. Comptuat. Imag. 3 (2017) 47–57.

[46] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, Proc. IEEE ICCV (2015) 1026–1034.

[47] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: graph, Models Image Process. 53 (1991) 231–239.

[48] K. Epperson, A.M. Sawyer, M. Lustig, M. Alley, M. Uecker, Creation of fully sampled MR data repository for compressed sensing of the knee, Proc SMRT (2013).

[49] S. Bakas, M. Reyes, A. Jakab, et al, Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, arXiv:1811.02629 (2018) https://arxiv.org/abs/1811.02629.

[50] J.H. Duyn, P.V. Gelderen, T.Q. Li, J.A.D. Zwart, A.P. Koretsky, M. Fukunaga, High-field MRI of brain cortical substructure based on signal phase, Proc. Natl. Acad. Sci. USA 104 (2007) 11796–11801.

[51] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, and M. Devin, TensorFlow: Large-scale machine learning on heterogeneous distributed systems, arXiv:1603.04467 (2016) https://arxiv.org/abs/1603.04467.

[52] D. Kingma, and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 (2014) https://arxiv.org/abs/1412.6980.

[53] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.

[54] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, Proc. ACM-MM (2014) 675–678.

[55] J.C. Ye, Y. Han, E. Cha, Deep convolutional framelets: a general deep learning framework for inverse problems, SIAM J. Imag. Sci. 11 (2018) 991–1048.

[56] M. A. Dedmari, S. Conjeti, S. Estrada, P. Ehses, T. Stöcker, and M. Reuter, Complex fully convolutional neural networks for MR image reconstruction, arXiv: 1807.03343 (2018) https://arxiv.org/abs/1807.03343.

[57] Y. Chen, F. Shi, A.G. Christodoulou, Y. Xie, Z. Zhou, D. Li, Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network, Proc. MICCA I (2018) 91–99.

[58] C. Pham, A. Ducournau, R. Fablet, F. Rousseau, Brain MRI super-resolution using deep 3D convolutional networks, Proc. IEEE ISB I (2017) 197–200.