



Original contribution

Compressed sensing MRI via a multi-scale dilated residual convolution network



Yuxiang Dai, Peixian Zhuang*

Jiangsu Key Laboratory of Meteorological Observation and Information Processing, Nanjing 210044, China

Jiangsu Technology and Engineering Center of Meteorological Sensor Network, Nanjing 210044, China

School of Electronic and Information Engineering, Nanjing 210044, China

Nanjing University of Information Science and Technology, Nanjing 210044, China

ARTICLE INFO

Keywords:

MRI reconstruction

Dilated convolution

Residual learning

Multi-scale

ABSTRACT

Magnetic resonance imaging (MRI) reconstruction is an active inverse problem which can be addressed by conventional compressed sensing (CS) MRI algorithms that exploit the sparse nature of MRI in an iterative optimization-based manner. However, two main drawbacks of iterative optimization-based CSMRI methods are time-consuming and are limited in model capacity. Meanwhile, one main challenge for recent deep learning-based CSMRI is the trade-off between model performance and network size. To address the above issues, we develop a new multi-scale dilated network for MRI reconstruction with high speed and outstanding performance. Comparing to convolutional kernels with same receptive fields, dilated convolutions reduce network parameters with smaller kernels and expand receptive fields of kernels to obtain almost same information. To maintain the abundance of features, we present global and local residual learnings to extract more image edges and details. Then we utilize concatenation layers to fuse multi-scale features and residual learnings for better reconstruction. Compared with several non-deep and deep learning CSMRI algorithms, the proposed method yields better reconstruction accuracy and noticeable visual improvements. In addition, we perform the noisy setting to verify the model stability, and then extend the proposed model on a MRI super-resolution task.

1. Introduction

MRI is a widely used imaging technology for visualizing the structure and functioning of the body with the advantages of non-radiation and non-ionizing nature. However, the slow imaging speed of MR poses a limitation on its widespread application. Recently, the CS theory [1] is introduced to reduce the MR scan time, and CSMRI can reconstruct a high resolution image from randomly sampled k -space data. The CSMRI problem can be formulated as the optimization

$$\hat{x} = \operatorname{argmin}_x \frac{1}{2} \|F_u x - y\|_2^2 + \sum_i \alpha_i R_i(x) \quad (1)$$

where x is denoted as the MRI to be reconstructed, y is the k -space data, and F_u represents the under-sampled Fourier encoding matrix. The first term $\|F_u x - y\|_2^2$ indicates data fidelity that can ensure the consistence between the Fourier coefficients of the reconstructed image and measured data. The second term R_i is an analytical, sparsifying transform term, and α_i is a factor for balancing data fidelity and transform terms. MR images can be generated by inverse Fourier transform of the

sampled k -space data, which are the Fourier coefficient of an object. However, aliasing artifacts (noise-like) are produced by the incoherence of under-sampled k -space in transform domain, as shown in Fig. 1.

To address this problem, a large number of CSMRI algorithms have been developed, and these methods tend to fall into two main categories: The first category of CSMRI algorithms are iterative optimization-based CSMRI, in which the sparsity is enforced in specific transform domain or underlying latent representation of images, and then an alternating iterative optimization scheme is adopted to CSMRI reconstruction [2–11]. A pioneering work of CSMRI is Sparse MRI [2], which exploits an off-the-shelf basis to capture a specific feature (wavelets recover point-like features, contourlets capture curve-like features). A hybrid TV regularizer combined with a L_0 -regularized tree-structured sparsity constraint [3] is introduced to overcome model-dependent deficiency and represent the measure of sparseness in wavelet domain. However, fixed bases fail to sparsely represent complicated MR images with underlying image edges and textures. To address this issue, several dictionary learning models (DLMRI [4], BPFA [5], and FDLCP [6]) and different wavelet regularizations based on

* Corresponding author at: Nanjing University of Information Science and Technology, Nanjing 210044, China.

E-mail address: zhuangpeixian0624@163.com (P. Zhuang).

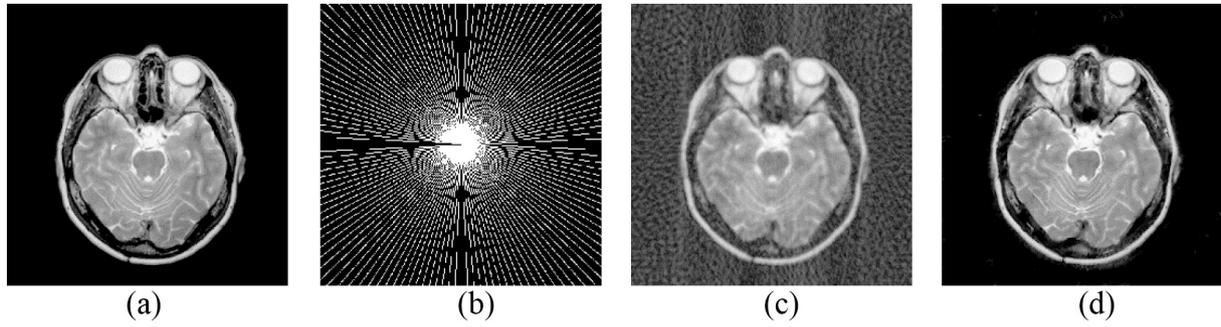


Fig. 1. The zero-filled reconstruction. (a) is a full-sampled MRI, (b) is a 20% radial sampling mask, (c) is the zero-filled reconstruction under (b), and (d) is the reconstruction using our method. Note that aliasing artifacts are clearly seen in the zero-filled reconstruction (c), which impair diagnostic information. However, our algorithm can remove these unpleasant artifacts (d).

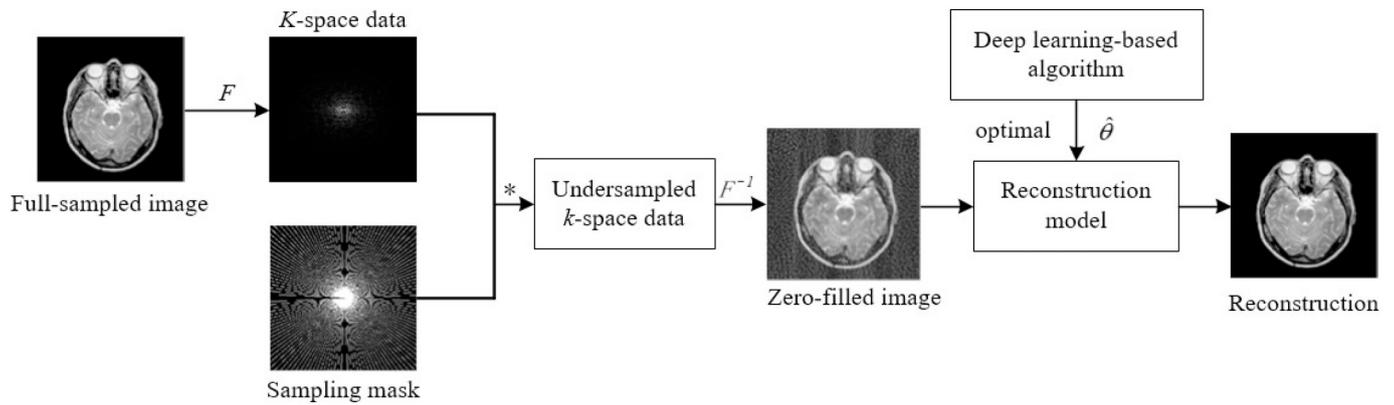


Fig. 2. The flowchart of the MR imaging based on deep learning approaches. F is Fourier transform, $*$ is the process of under-sampling on k -space data and F^{-1} represents inverse Fourier transform.

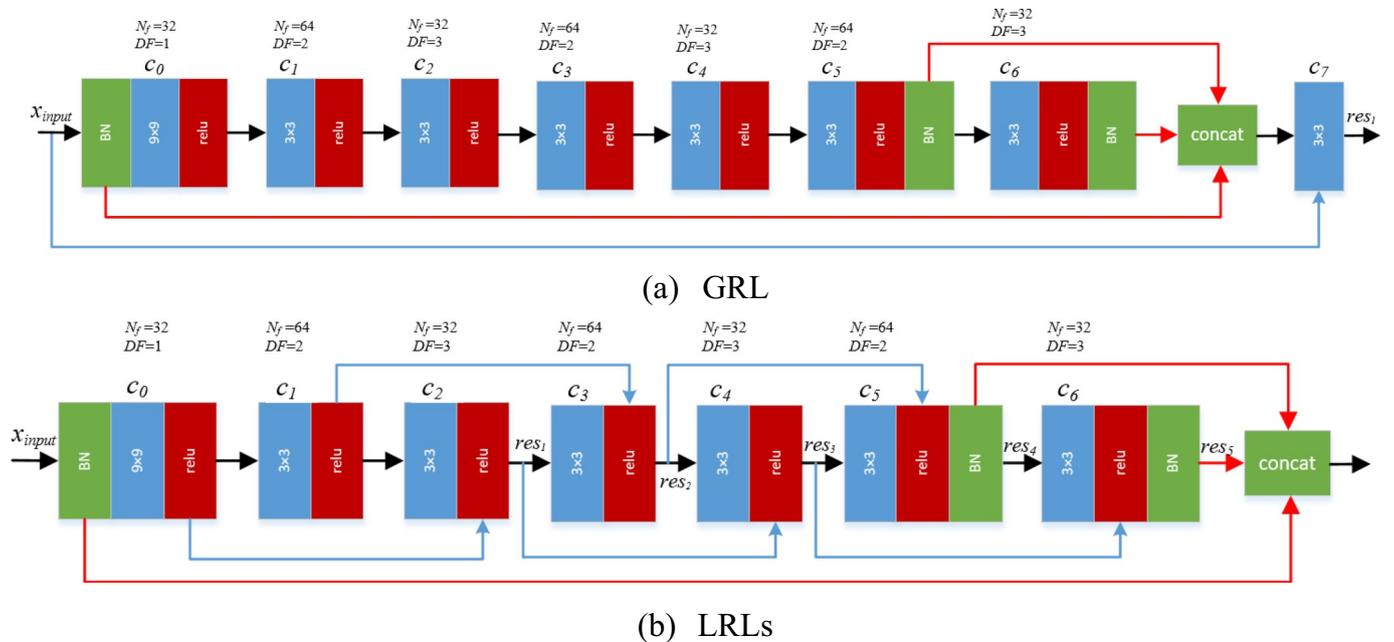


Fig. 3. GRL and LRLs based on dilated network. (a) is the GRL, which takes advantage of initial information. (b) is the LRLs, which further improves the information flow. Then MDN integrates both of them.

geometric information (PBDW [7] and PBDW with pFISTA [8]) are exploited. For instance, a fast orthogonal dictionary learning method (FDLCP) is introduced to provide adaptive sparse representation of images, in which image is divided into classified patches according to the same geometrical direction and dictionary is trained within each

class for enhanced sparsity. And patch-based directional wavelets model (PBDW) is proposed to promote MRI reconstruction, and patch geometric direction is trained from the reconstructed image using conventional CSMRI methods. But these dictionary learning or wavelet regularizations algorithms are required that parameters such as

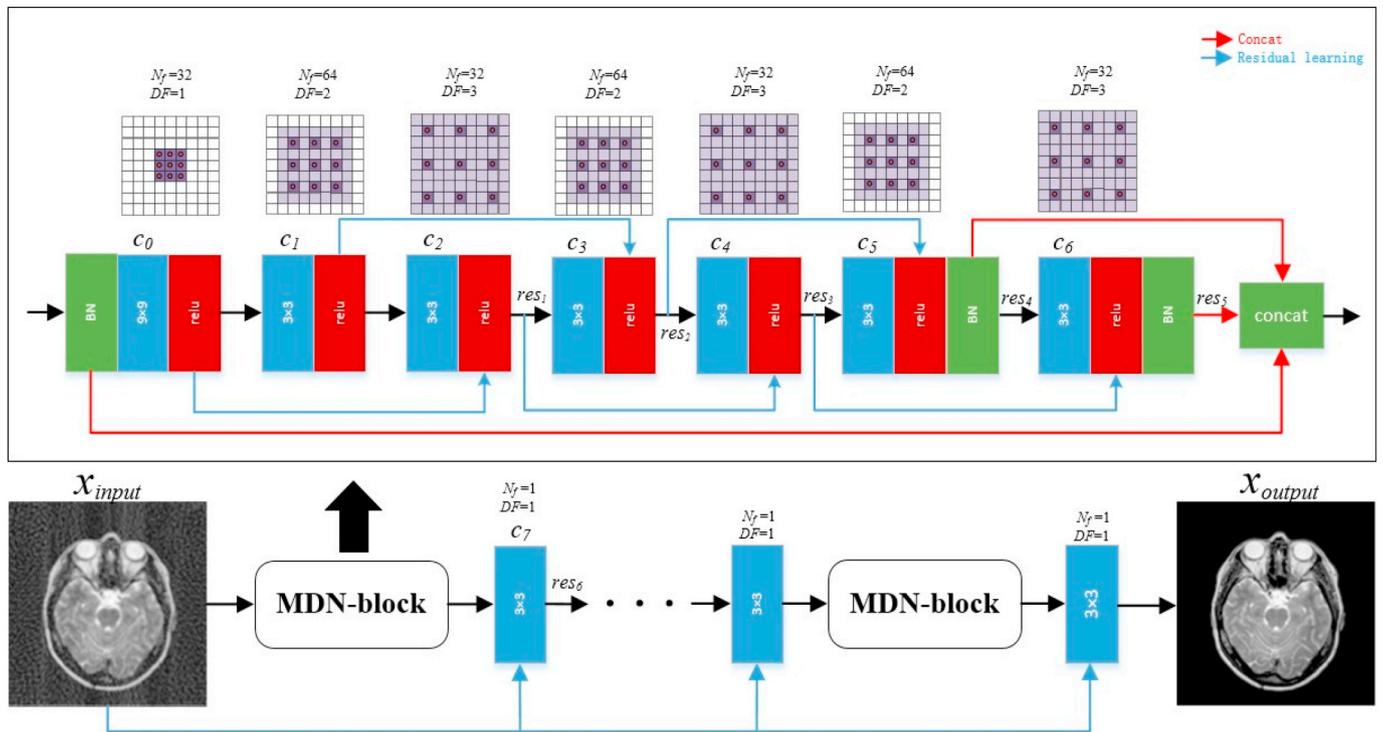


Fig. 4. Schema for the proposed MDN for CSMRI. The light purple regions stand for the receptive fields of the dilated convolutional kernels with different DF values. The blue lines are the input of residual learnings, and the red ones are the input of a concatenation layer. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

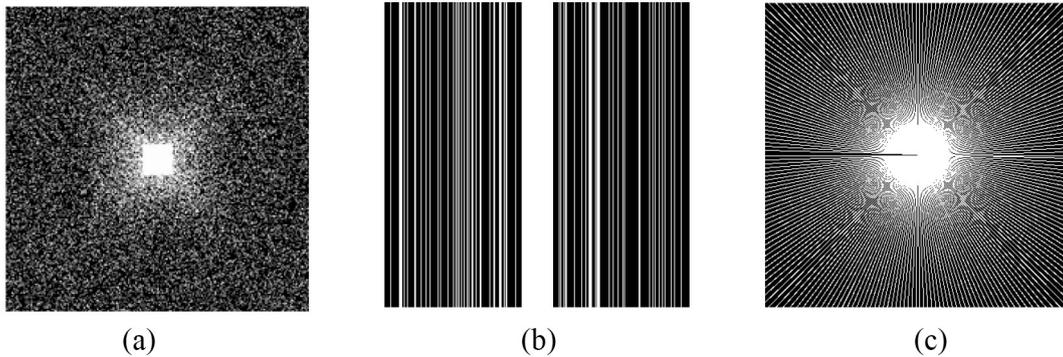


Fig. 5. Three sampling masks with specific sampling rates. (a) 20% variable density random sampling. (b) 25% cartesian sampling. (c) 30% radial sampling.

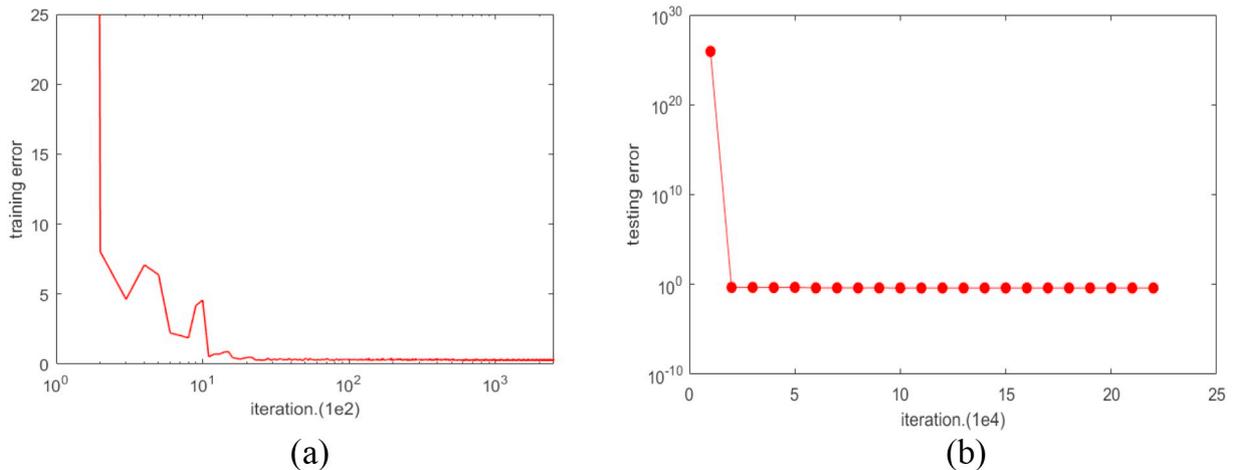


Fig. 6. The convergence curves of our network. (a) Training convergence curve. (b) Testing convergence curve.

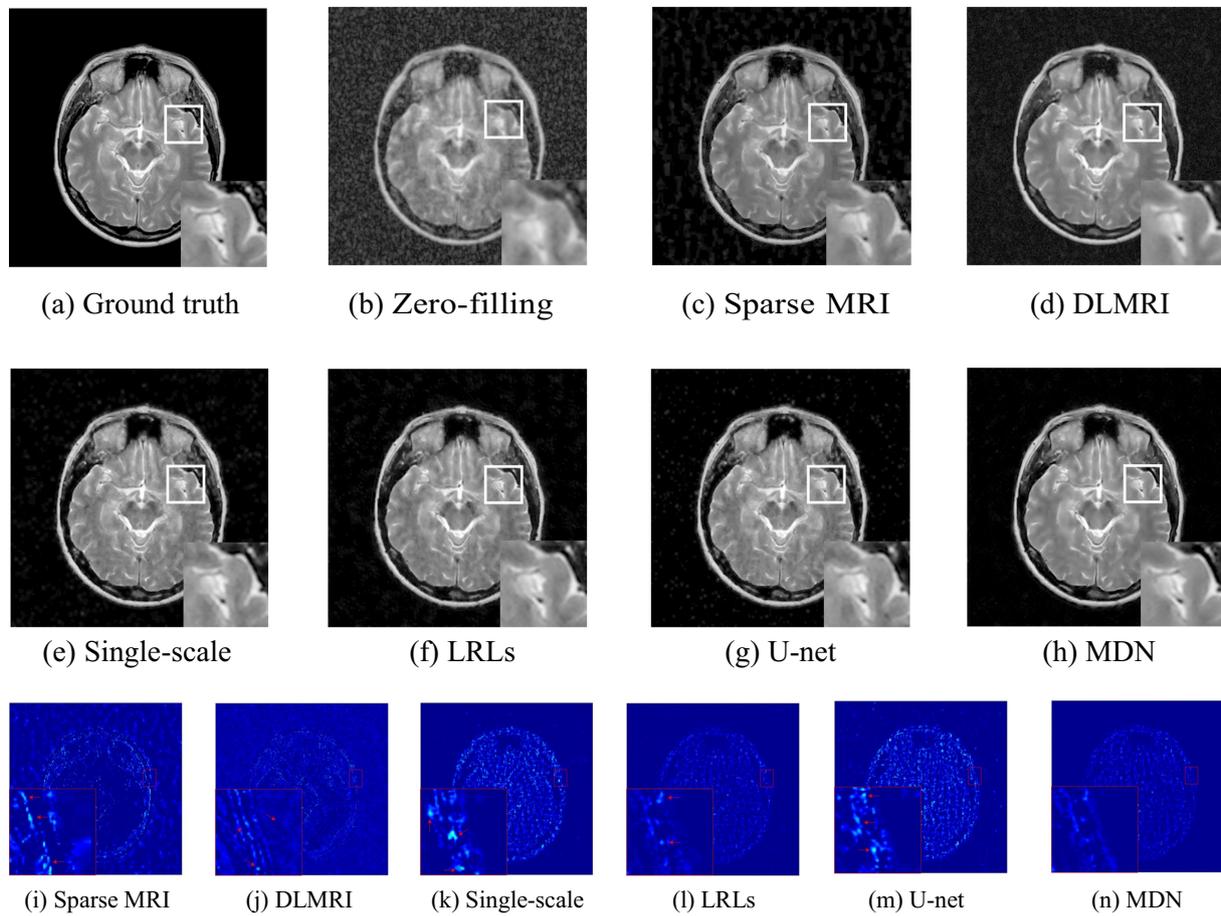


Fig. 7. Reconstruction results for 20% variable density sampling. (a) Original. (b)–(h) Reconstructed images. (i)–(n) The errors of six CSMRI methods.

dictionary size and patch sparsity are preset. A Bayesian non-parametric dictionary learning model (BPTV) [9] applies the beta process to learn the sparse representation necessary for CSMRI, in which beta process is an effective prior for non-parametric learning of dictionary parameters such as dictionary size and patch sparsity. In addition, some methods are performed to obtain the information from the MRI of interest. A method (PANO) [10] exploits nonlocal similarity of image patches by establishing a patch-based nonlocal operator, which effectively produces sparse vectors by operating on grouped similar patches of the image. Another MRI reconstruction algorithm can promote structures and suppress artifacts with an edge-preserving filtering prior [11], in which a gradient domain guided image filtering (GFF) is embedded. However, conventional CSMRI methods are limited in model capacity to recover diverse image structures, and require a lot of iterative operations which is time-consuming and fails in real-time reconstructions.

The deep learning-based CSMRI [12–16] can learn a nonlinear mapping from the zero-filled MRI to the fully-sampled MRI. In addition, better MR images can be reconstructed by exploiting existing training model with no additional iterations, which can achieve real-time execution compared with iterative optimization-based CSMRI. For the purpose of accelerating MR imaging, an off-line convolutional neural network (CNN) [12] is applied for CSMRI by learning an end-to-end mapping between zero-filled and fully-sampled MR images for the first time. After that, a deep cascade of CNNs [13] combines convolution and data sharing approaches to identify spatio-temporal correlations in MR images, which can boost data acquisition. In order to accelerate MR acquisition process with performance guarantee, U-net with deep residual learning [14] is proposed to formulate a CS problem as a residual regression problem where aliasing artifacts from under-sampled data

are simpler than those of images in structure. ADMM-Net [15] uses alternating direction method of multipliers (ADMM) to derive and define the data flow, which can optimize a general CSMRI model to reconstruct MR images from a small number of under-sampled data in k -space. Moreover, in the Bayesian deep learning model [16], the MC-dropout and heteroscedastic loss are applied to the reconstruction networks to model epistemic and aleatoric uncertainty which can achieve competitive performance. Although the above-mentioned deep learning algorithms can accelerate MR acquisition process with performance guarantee, they are composed of complex network with more parameters.

To address these limitations, we develop a novel multi-scale dilated network (MDN) for MRI reconstruction. The contributions of our paper are summarized as follows:

- We develop a dilated network to expand the receptive field of convolutional kernels for reducing network parameters without the loss of resolution, which can obtain multi-scale information. When compared with the larger kernels with same receptive field, the dilated network can increase reconstruction accuracy and accelerate training speed.
- Considering the structures and details synthetically, we adopt global residual learning to make up the overall structural features missed during extracting process, and employ local residual learnings to extract more abundant features to preserve better edges and details.
- We exploit concatenation layers to fuse multi-scale features, which can make full use of the abundance of features to maintain image details for better reconstruction results.
- We perform numerous experiments to demonstrate the better capability of the proposed model with three sampling masks and a

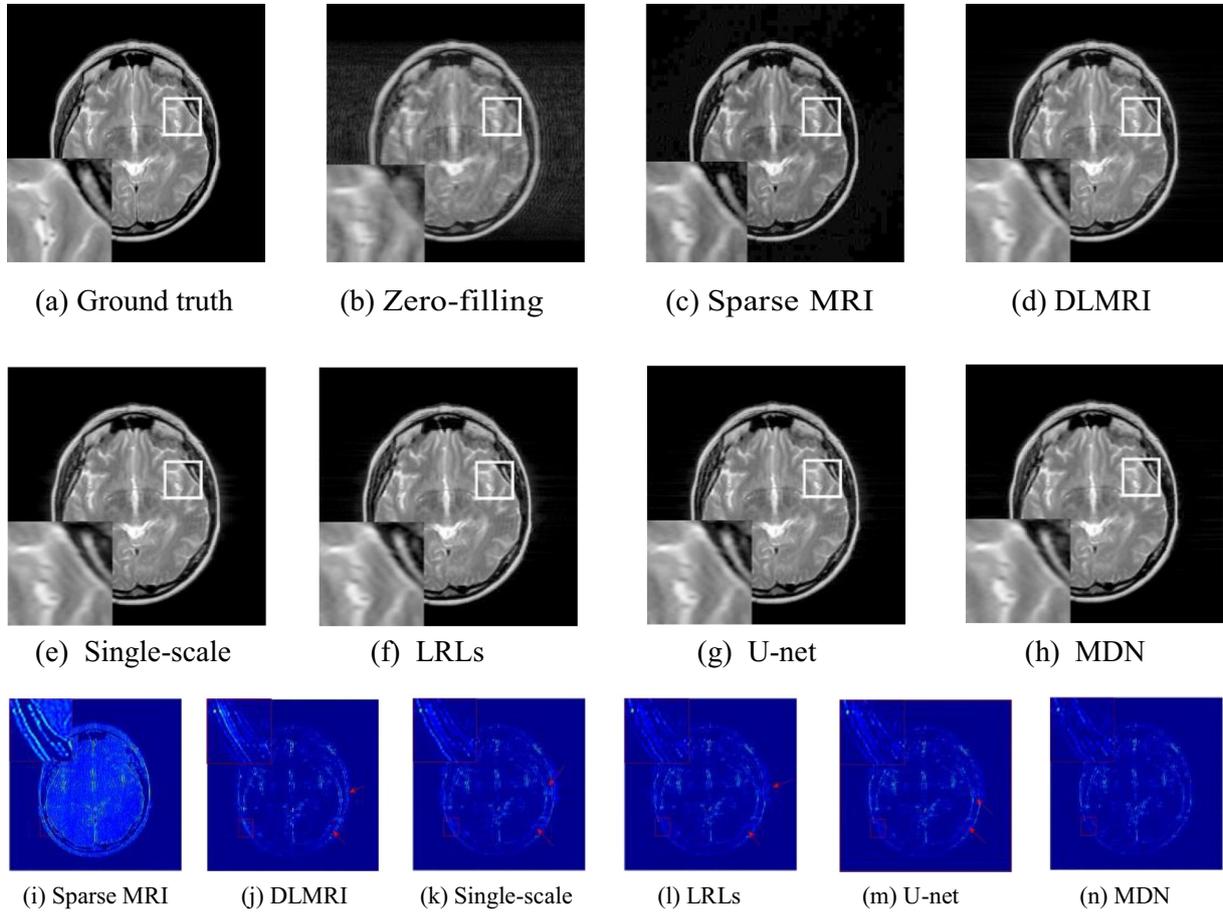


Fig. 8. Reconstruction results for 25% cartesian sampling. (a) Original. (b)–(h) Reconstructed images. (i)–(n) The errors of six CSMRI methods.

variety of sampling rates for each mask. In addition, the proposed model can be applied into MRI noisy setting and super-resolution tasks, which demonstrate the effectiveness of the proposed model.

2. Related work

In this section, we review the related components of deep learning that are used in the proposed network for MRI reconstruction.

2.1. Residual learning

As the number of network layer increased, the expression of the overall model is enhanced, which results in poor training accuracy. The deep residual network [17] introduces an equal fast connection to solve the problem of gradient disappearance. The basic block of residual is to use a shortcut during two contiguous convolutional layers. Residual learning has achieved impressive performance on image low-level tasks, such as reconstruction [18–21], super-resolution [22–24], denoising [25–27], deraining [28–30], etc.

2.2. Dilated convolution

Dilated convolution (conv) has been proposed for intensive prediction tasks [31–33]. Compared with the ordinary convolution, dilated convolution has a dilated rate parameter called dilated factor (DF), which is mainly used to indicate the dilatation size. Dilated convolution has a larger receptive field, while keeping the number of kernel parameters constant with the same size as the ordinary convolution. The feature map size of the output can be stayed the same by dilated convolution.

2.3. Concatenation

The concatenation (concat) layer [34–36] is used to splice two or more feature maps in the channel or number dimension without operating the residual layer, which can fuse single-scale or multi-scale features. For instance, when conv_1 and conv_2 are spliced on the channel dimension (k_1, k_2), the other dimensions (N_i, H , and W) must be consistent, where N_i is the number of image patches, H and W represent the height and width of output matrix respectively. The operation at this time is channel k_1 plus channel k_2 and the output of the concat layer can be expressed as: $N_i \times (k_1 + k_2) \times H \times W$. The concat layer is generally used to employ the semantic information of multi-scale feature maps to achieve better performance.

3. Method

3.1. Problem formulation

Different from traditional CSMRI problem, deep learning-based algorithms can generate a pre-trained model which can be directly transferred in MR imaging. And the deep learning-based algorithms require to train huge measurements for the model with optimal performance, which can be seen in Fig. 2. The deep learning-based CSMRI problem can be formulated as:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \frac{1}{2} \|F_u x - y\|_2^2 + \xi \|x - f_{\text{cnn}}(x_u | \hat{\theta})\|_2^2 \quad (2)$$

where f_{cnn} is the forward propagation of the CNN with the parameter θ that contains millions of network weights, and ξ is a regularization parameter. x_u represents the zero-filled images as shown in Fig. 2, and $\hat{\theta}$

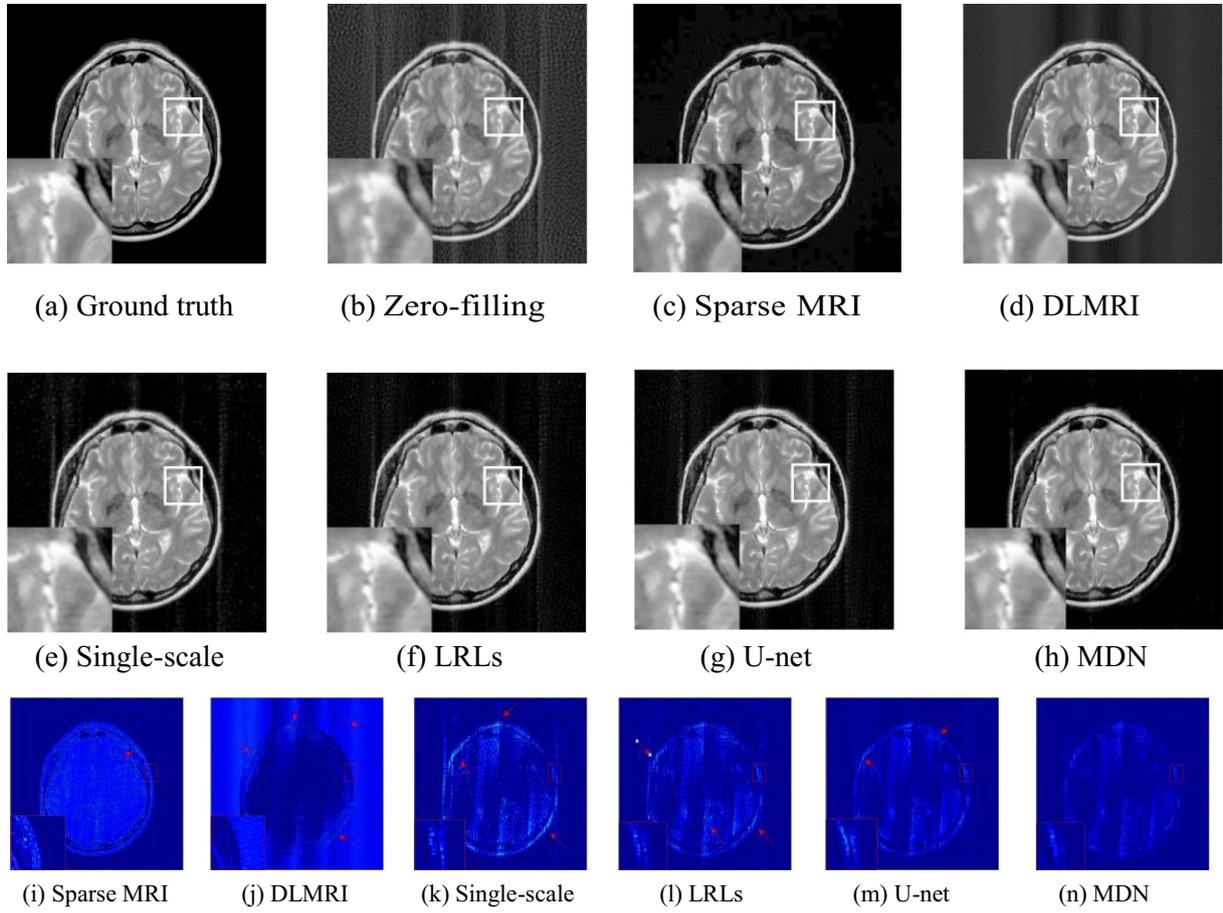


Fig. 9. Reconstruction results for 30% radial sampling. (a) Original. (b)–(h) Reconstructed images. (i)–(n) The errors of six CSMRI methods.

are the optimal parameters of trained CNN.

3.2. Proposed block

A multi-scale dilated network (MDN) block consists of dilated convolution with rectified linear units (relu) [37], residual learning and multi-scale concatenation. Fig. 4 shows the overall framework of the proposed MDN architecture. Then we present the compositions of the proposed block in details.

3.2.1. Dilated convolution

As shown in Fig. 4, there are 7 convolutional layers in a MDN-block, which consists of one normal convolution, three 2-dilated convolutions and three 3-dilated convolutions. As we all know, a convolutional layer can extract n layers of features when the N_f is set to n , in which N_f is the number of filters (convolutional kernels) in the convolutional layers. It is well-known that more features are extracted as the number of kernels becomes larger, and the effect of network training increases accordingly. For reducing the parameters to lower the computational complexity, we choose proper DF and N_f for convolutional layers. In Fig. 4, DF is set to 3 when N_f is 32, conversely, DF is set to 2 when N_f is 64.

We increase the receptive field of the convolutional kernel to expand the receiving domain of image information. Since the number of convolutional kernels is limited, appropriate DF should be chosen to match the proposed network and data sets. Compared with the original convolutional layer (the size of kernel is increased), the dilated convolution achieves comparable performance with less parameters demonstrated in experiments. The key is to obtain a better tradeoff among the number, size and dilated factor of convolutional kernels.

3.2.2. Local and global residual learnings

The global and local residual learnings are integrated to maintain the abundance of feature maps for better reconstruction. Global residual learning (GRL) tries to obtain initial information, while local residual learnings (LRLs) are utilized to further improve the information flow. Fig. 3(a) shows that the GRL concatenates a series of convolution layers, and finally connects the input to the output for preventing the loss of features. Fig. 3(b) presents the LRLs where there are five local residuals in a block, which do not burden the network complexity. Fig. 4 is the proposed residual network. We combine GRL and LRLs to ensure adequate features, slightly increase the network complexity, and achieve better reconstruction results than the above two residual learnings.

GRL:

$$c_n = f(c_{n-1}), res_1 = I_{input} + c_7 \quad (3)$$

LRLs:

$$res_1 = c_0 + c_2, res_2 = c_1 + c_3, res_3 = res_1 + c_4, res_4 = res_2 + c_5, res_5 = res_3 + c_6. \quad (4)$$

Proposed residual:

$$res_1 = c_0 + c_2, res_2 = c_1 + c_3, res_3 = res_1 + c_4, res_4 = res_2 + c_5, res_5 = res_3 + c_6, res_6 = I_{input} + c_7 \quad (5)$$

where $f(\cdot)$ represents the operation of convolutional layer and activation function, c_n denotes feature maps of the n -th convolutional

Table 1

PSNR/SSIM of different methods on real-valued brain MRI as a function of sampling percentage. Sampling masks include cartesian sampling, variable density random sampling and radial sampling. The value with red bold font indicates ranking the first place while value with blue font is the second place.

Mask	Sampling%	Sparse MRI	DLMRI	Single-scale	LRLs	U-net	MDN
Cartesian	10	24.50/0.811	25.22/0.726	25.46/0.797	26.14/0.802	26.15/0.815	26.59/0.840
	15	26.16/0.857	28.37/0.841	28.29/0.861	28.62/0.860	28.29/0.850	28.86/0.871
	20	26.98/0.885	30.68/0.902	30.53/0.905	31.05/0.907	30.80/0.907	31.43/0.930
	25	27.60/0.895	32.85/0.934	32.30/0.930	32.47/0.931	33.13/0.939	33.25/0.950
	30	28.45/0.892	34.77/0.955	34.28/0.954	34.57/0.954	34.81/0.958	35.27/0.967
Random	10	27.38/0.776	31.27/0.554	32.07/0.904	31.51/0.887	32.01/0.902	32.16/0.913
	15	27.64/0.821	32.86/0.612	32.76/0.908	33.15/0.876	33.22/0.920	33.82/0.930
	20	30.44/0.888	34.34/0.675	33.99/0.924	34.17/0.919	34.70/0.942	34.95/0.944
	25	33.44/0.915	35.75/0.727	34.97/0.939	35.02/0.928	35.77/0.947	35.96/0.948
	30	34.71/0.963	36.75/0.754	35.86/0.949	35.74/0.936	36.63/0.954	36.83/0.958
Radial	10	23.17/0.668	27.93/0.405	28.19/0.815	28.98/0.844	29.00/0.849	29.64/0.873
	15	24.68/0.742	29.77/0.448	30.26/0.860	30.96/0.877	30.67/0.871	31.87/0.905
	20	25.91/0.648	30.55/0.467	31.97/0.888	32.54/0.888	32.48/0.889	33.48/0.925
	25	26.14/0.773	31.02/0.478	33.21/0.917	33.75/0.921	33.84/0.925	34.51/0.938
	30	28.26/0.898	31.35/0.487	34.21/0.935	34.91/0.928	35.12/0.932	35.64/0.955

layer, res_n is the n -th residual sum, and I_{input} represents the input images. The proposed residual learnings have a better effect on MRI reconstruction without burdening the network complexity. We utilize both GRL and LRLs in the network to prevent the loss of valid features.

3.2.3. Multi-scale concatenation

The computational complexity of residual block with dilated convolutions shows a very high growth trend, especially in the case of huge data sets. To solve the shortcoming, we exploit a multi-scale residual block, in which different numbers of convolutional kernels and residual learnings are integrated to enrich features. At the same time, multi-scale features are stacked so that abundant information can be shared and reused, which contributes to the fusion of local features. In addition, the application of a 3×3 kernel after concat (that is after a block) aims to facilitate the fusion of features and cut down computational complexity, and batch normalization [38] is utilized before the input of concat to improve accuracy and accelerate convergence, which can accelerate MR imaging.

3.3. Network architecture

As discussed above, the proposed MDN framework consists of repeated blocks. The residual sum after a block aims to supplement initial information missing in the process of extracting features during the previous blocks (Fig. 4). However, deeper the network with more blocks does not mean that the extracted features are more favorable to reconstruction results. Proper number of blocks should be adopted for CSMRI.

The loss function of the proposed network is:

$$Loss = \frac{1}{2M} \sum_{i=1}^M (\hat{x}_i - x_i)^2 \quad (6)$$

where x_i denotes the full-sampled image, and \hat{x}_i represents the output of the network; M is the number of training images. The proposed network can be implemented using Caffe, Pytorch or Tensorflow.

4. Experiment results

We provide numerous experiments to demonstrate the effectiveness of the proposed method in MR reconstruction, which compares with several iterative optimization-based and deep learning-based approaches. We employ three sampling masks: variable density sampling [2], cartesian sampling [39] and radial sampling [40], and a variety of sampling rates are set for each mask. An example of each mask is shown in Fig. 5. Then we consider the noisy settings and apply the proposed model into MR super-resolution. In addition, the ablation study on residual learnings is conducted to illustrate the effect of GRL and LRLs, and different initial learning rates are considered in experiments.

4.1. Implementation details

We train and test the network based on the NVIDIA GeForce GTX1080Ti with 11GB GPU memory. We use Caffe for algorithm in network training and Matlab for preprocessing of data sets. The maximum iterations of the network are 250,000. The main function of the solver in Caffe is to alternately transfer forward and backward conduction to update the weight of neural network, so as to minimize the loss. The optimization we used is Adam, the base learning rate is set to 0.001, and the weight decay is set to 0.0001, which is weight attenuation term used to prevent over-fitting. The learning rate policy is

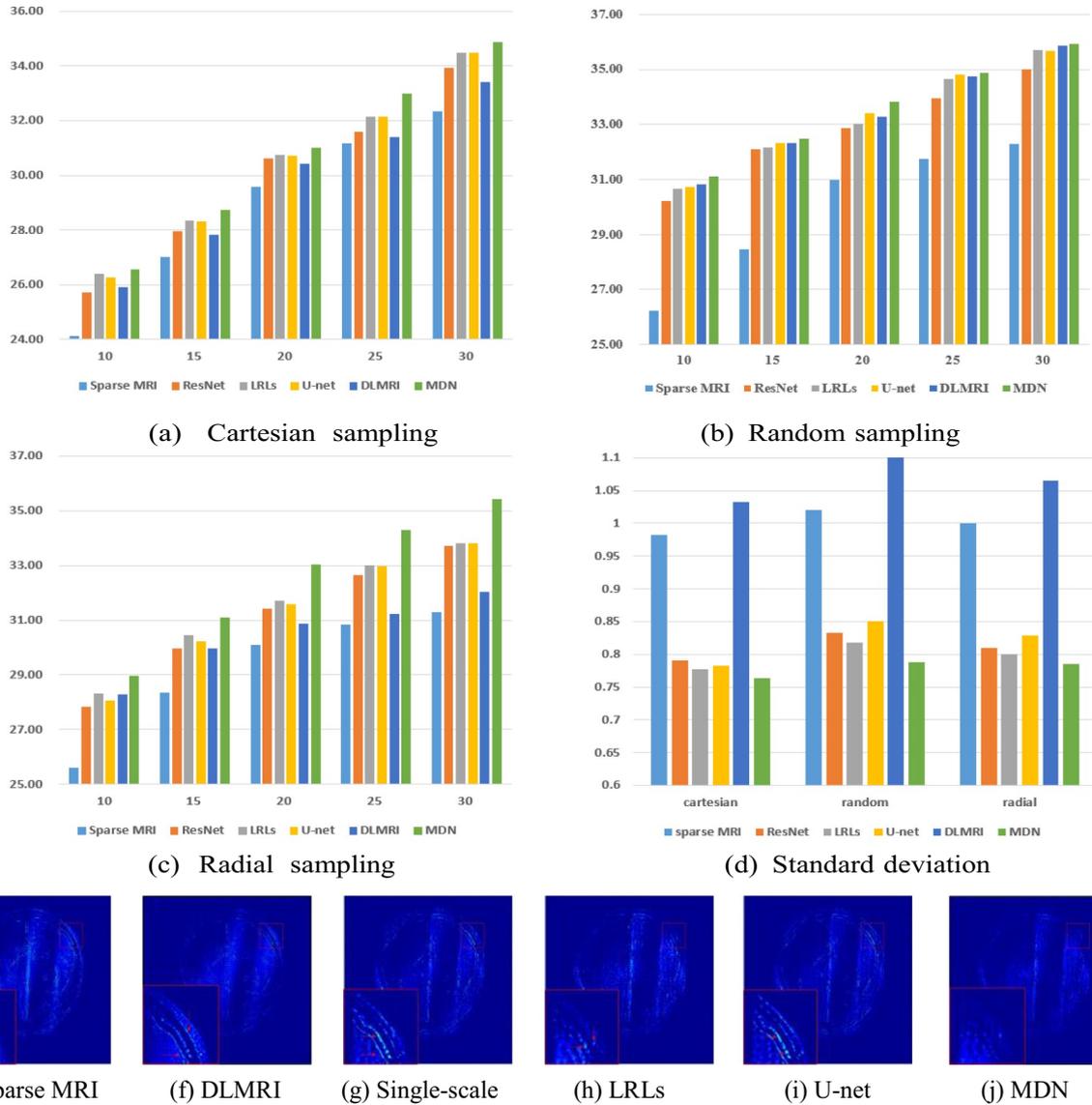


Fig. 10. The reconstruction results on complex-valued MRI data. (a)-(c) are the PSNR values of different methods with three sampling masks and five sampling rates, in which the x-axis represents sampling rates. (d) is the standard deviation of PSNR values on different methods when using 30% sampling rates, in which the x-axis represents sampling masks. And (e)-(j) are the errors of six CSMRI methods.

Table 2

PSNR/SSIM/training time for the different combinations of N_f and DF from 2-nd layer to 7-th layer. The first layer is fixed: $N_f = 32$, $DF = 1$. The value with red bold font indicates ranking the first place in this column while values with blue font are the second and third place.

N_f	DF	PSNR	SSIM	Training time (min)
64×6	3-3-3-3-3-3	34.64	0.946	782.5
	2-2-2-2-2-2	34.98	0.940	720
	2-3-2-3-2-3	34.88	0.945	752.5
	3-2-3-2-3-2	34.97	0.937	752.5
32×6	3-3-3-3-3-3	34.85	0.940	685
	2-2-2-2-2-2	34.62	0.931	645
	2-3-2-3-2-3	34.83	0.930	662.5
	3-2-3-2-3-2	34.83	0.931	667.5
64-32-64-32-64-32	2-3-2-3-2-3	34.95	0.944	700

Table 3

PSNR/SSIM for the ablation study on concatenation layers with different sampling masks and specific sampling rates.

Mask & sampling rate	Random & 20%	Cartesian & 25%	Radial & 30%
No concat	33.54/0.903	32.49/0.923	34.41/0.872
With concat	34.95/0.944	33.25/0.950	35.64/0.955

Table 4

PSNR/SSIM for the ablation study about residual learnings based on dilated convolutions.

Learning rate	0.0001	0.001	0.01
GRL ✓ LRLs ✓	34.95/0.935	34.95/0.944	34.31/0.939
GRL × LRLs ✓	34.15/0.912	34.17/0.919	33.74/0.924
GRL ✓ LRLs ×	34.31/0.905	34.52/0.902	31.61/0.662
GRL × LRLs ×	32.00/0.781	31.61/0.702	31.61/0.702

Table 5
PSNR for 35% variable density sampling of brain MR with various noise standard deviations.

Reconstruction method	$v = 0$	$v = 0.01$	$v = 0.02$	$v = 0.03$
Zero-filled & noisy	31.50	20.88	18.34	16.81
LRLs	37.53	31.09	29.78	29.22
MDN	38.06	31.74	30.36	29.54

step and the gamma coefficient associated with the learning rate is set to 0.1. The step size indicates the frequency at which we should go to the next training step, which is set to 50,000. The weight of the last gradient update (momentum) is set to 0.9. The training errors are displayed per 100 iterations and testing errors are displayed per epoch, which can be seen in Fig. 6.

4.2. Data sets

Our real-valued data sets come from the MRI Multiple Sclerosis Database (MRI MS DB).¹ Among them, we select 450 T2-MR images as a training set. In addition, we expand this training set to 1534 images by rotating these 450 pictures, and we consider 50 high quality T2 images as a testing set. Moreover, we choose 800 simulated complex images as a training set, and select 80 simulated complex images as a testing set. All images have the size of 378×378 .

4.3. Metrics

We not only evaluate the reconstructed results subjectively, but also use two objective evaluation indicators: peak signal to noise ratio (PSNR) [41] and structural similarity index (SSIM) [42]. The PSNR represents the ratio between the power of the maximum possible image intensity across a volume and the power of distorting noise and other errors, and the SSIM shows the similarity between two images by exploiting the inter-dependencies among nearby pixels. Higher values of PSNR and SSIM demonstrate better reconstruction. Additionally, we employ the standard deviation of PSNR to demonstrate the network stability on complex-valued data.

4.4. Quantitative evaluation

To evaluate the reconstruction performance, we compare the proposed model with the two iterative optimization-based methods: Sparse MRI [2] and DLMRI [4], and three deep-learning algorithms: Single-scale residual learning (Single-scale) [14], LRLs and U-net [14]. The former two optimizations are provided by the authors' homepage. The latter three deep learning algorithms are reproduced using Caffe. We consider the zero-filled reconstruction results as well. We reproduce Single-scale residual learning in the same environment, which uses a modified deconvolution network with symmetric contracting path. Based on Single-scale residual learning, the U-net utilizes the pooling layer and deconvolution to make full use of multi-scale features. LRLs have been shown in Fig. 3(b).

4.5. Experiments on real-valued MRI with different masks

As shown in Figs. 7, 8 and 9, Sparse MRI and DLMRI have a lot of unpleasant artifacts, Residual learning and U-net can eliminate most of artifacts, but are not ideal for restoring image details. However, the proposed method can reconstruct better MR images, which outperforms other competitive methods in visualization of structures reconstruction

¹ <http://www.medinfo.cs.ucy.ac.cy/index.php/facilities/32-software/218-datasets>.

and artifacts removal. Meanwhile, we can see from the absolute error residuals for three sampling experiments that the proposed MDN algorithm restores a finer detail structure than other algorithms. Moreover, we present the PSNR and SSIM values in Table 1 for different algorithms, sampling masks and sampling rates. It is demonstrated that the proposed method provides better reconstruction performance and visual results than other competitive methods. We can also see the obvious improvement of all algorithms over zero-filling both in visualization and objective metrics. In particular, a higher SSIM value of Sparse MRI appears when using 30% variable density random sampling, however, Sparse MRI generates more artifacts than the proposed MDN.

4.6. Experiments on complex-valued MRI with different masks

We evaluate the performance of the proposed model using PSNR on complex-valued data and compare with two optimization-based methods and three deep-learning methods. We present the PSNR results for all sampling masks and five rates in Fig. 10(a)–(c) and it is obvious that the proposed model outperforms other five methods, which can demonstrate the effectiveness of MDN model on complex-valued data. Additionally, we provide the standard deviation on 80 test images of different methods when using 30% sampling rates of three masks in Fig. 10(d). We can observe that deep-learning methods obtain more stable performance than DLMRI and Sparse MRI. In Fig. 10(e)–(j), we show the absolute value of residuals of different algorithms using 30% radial sampling rate. We can see that the proposed model has less noise-like errors than other five methods.

4.7. Ablation study

4.7.1. Ablation study on network size setting

As mentioned above, we choose proper DF and N_f under the consideration of network size and performance. We conduct several experiments about the setting of DF and N_f in Table 2 and demonstrate the PSNR/SSIM values of different combinations. Additionally, we show the training time to evaluate computational cost with various network sizes.

In MDN blocks, the first layer with 9×9 kernel aims to enlarge receptive fields to extract more initial information for the block with no necessary to employ larger DF and N_f . We make a comparison between 9×9 kernel with 32 filters and 3×3 kernel with 64 filters in the first layer, and the former increases the value of PSNR by 0.1 than the latter. Therefore, we fix the first layer as shown in Fig. 4. In Table 2, all channels (N_f) of feature maps in MDN blocks set to 64 indeed increases the training time with a little improvement in PSNR/SSIM, however, all N_f set to 32 decreases reconstruction results in despite of less training time. Considering training time, reconstruction results and application of local residual learnings, we choose the alternating N_f values of 64 and 32. Meanwhile, we employ larger DF values for the layers with 32 feature maps in order to supplement some useful information extracted by enlarged receptive fields. By the way, setting larger DF than 3 obviously burdens the network and increases the training time.

4.7.2. Ablation study on the concat layer

To demonstrate the effectiveness of fusing multi-scale features, we conduct the ablation investigation on concatenation layers. It can be noticeable in Table 3 that using concat layers to fuse multi-scale features extracted from dilated network can achieve better reconstruction.

4.7.3. Ablation study on residual learnings and investigation on initial learning rates

We have explained that the proposed MDN integrates GRL and LRLs to maintain the abundance of feature maps for better reconstruction. And in this section, we show the results in terms of PSNR and SSIM among non-residual, global residual, local residual and MDN, in which all of them are based on multi-scale dilated network. As shown in

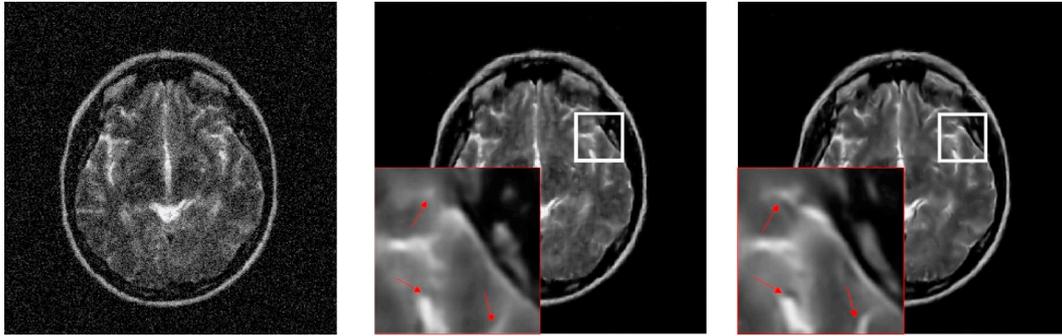


Fig. 11. Reconstruction for noisy images based on zero-filling, in which the noise standard deviation is set to 0.01.

Table 6
PSNR/SSIM for non-dilated and dilated networks with several blocks.

Block	1	2	3
Non-dilated	34.68/0.917	34.78/0.940	33.95/0.919
Dilated	34.85/0.930	34.95/0.944	34.98/0.934

Table 4, MDN which integrates GRL and LRLs outperforms other residual learnings and non-residual learning. It is obvious that MDN extracts more valid feature maps which can provide better reconstruction. Based on residual experiments, we consider the effort of different initial learning rates on reconstruction as well. It can be noticed in Table 4 that the four networks generally perform outstandingly in 0.001. As a consequence, we set initial learning rate as 0.001 during all training process.

4.8. Experiments in the noisy setting

The MR imaging we considered above have been completely noiseless. However, unexpected noise may be mixed in the sampling process for some external conditions. We continue our evaluation of noisy MRI to verify the stability of reconstruction based on MDN. Moreover, we compare the proposed MDN with one deep learning-based algorithm (LRLs) in terms of visualization and metrics. The noisy MR images are respectively mixed with complex white Gaussian noise having standard deviation $v = 0.01, 0.02, 0.03$. And ground truth is the original noise-free MRI. It can be noted that the proposed MDN achieves better results than the LRLs method in terms of PSNR in Table 5. Fig. 11 shows the reconstruction based on MDN, in which the noisy image has been well recovered.

Table 7
PSNR/SSIM for scale factors $\times 2, \times 3, \times 4$ on our data sets.

Scale	2	3	4
MDN	38.73/0.986	34.06/0.962	30.09/0.917
VDSR	38.04/0.983	30.85/0.930	29.60/0.906
difference	0.69/0.003	3.21/0.032	0.49/0.011

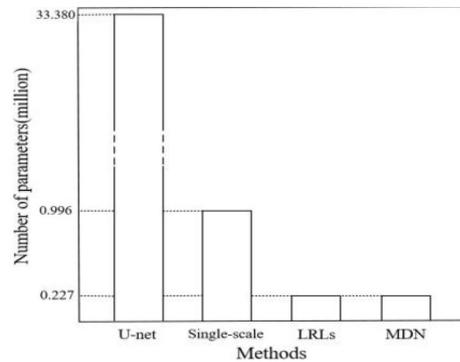
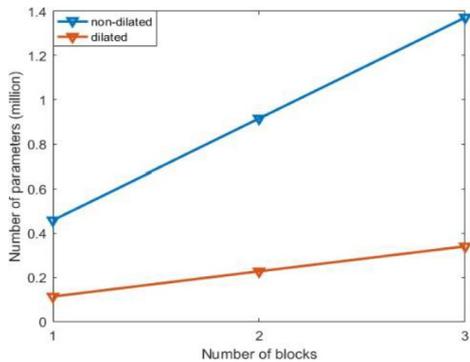
4.9. Discussions on dilated convolutions, the number of blocks and parameters

We also verify the effect of the number of MDN-blocks, and calculate the number of correspond parameters, which aims to obtain a tradeoff between network size and performance. For non-dilated network, we control the receptive fields of convolutions consistent with dilated convolution. For the receptive field, 2-dilated & 3×3 convolution is equivalent to non-dilated & 5×5 convolution; and 3-dilated & 3×3 convolution is equivalent to non-dilated & 7×7 convolution.

From the results of Table 6 and Fig. 12 referring to parameters calculation, two dilated blocks perform better reconstruction with less parameters. And it is obvious that the MDN achieves better reconstruction results with least parameters than other deep learning methods. As a consequence, the number of blocks should be set to 2, which can perform better results with a guarantee of training speed for the huge data sets.

4.10. Experiments on super-resolution

Subsequently, we conduct extended experiments on MR image super-resolution, which aims to recover high-resolution MR images from their low-resolution images for improving image analysis and visualization in the clinic. VDSR [24] trains a deep network with multiple



(a) Number of blocks vs number of parameters. (b) Deep learning methods vs number of parameters

Fig. 12. The number of parameters.

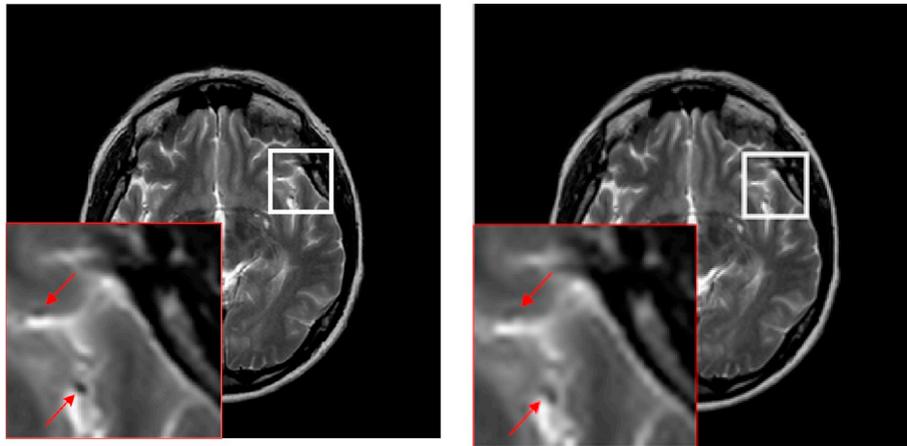


Fig. 13. Super-resolution results of a brain MRI with scale factor $\times 2$.

scale factors for image super-resolution task which can reduce the number of parameters and achieve efficient results. We demonstrate the comparison results of the proposed MDN and VDSR in Table 7 and Fig. 13. It is noted that the proposed MDN performs better reconstruction results than VDSR on a huge data set.

5. Conclusion and prospect

A novel multi-scale dilated network (MDN) has been presented for CSMRI. The proposed MDN is composed by cascading two basic blocks where dilated convolutions, global and local residual learnings, and concatenation layers are integrated to extend the receptive fields of convolutional kernels for reducing network parameters, maintaining features abundance, and fusing multi-scale features, respectively. Final experiments demonstrate that MDN achieves outstanding performance with training huge and diverse data, and the proposed network outperforms several competitive CSMRI algorithms in subjective and objective assessments. In addition, the proposed model is effective in MR noisy setting and super-resolution tasks.

In the future, we will adjust our model to parallel and dynamic imaging referred from [13,43]. And we will also improve our method with some variational models [44,45] which is beneficial to image reconstruction. In addition to MR reconstruction, we will consider the application of our model in segmentation task [21].

Acknowledgement

The authors sincerely thank anonymous editor and reviewers for their constructive and valuable comments. This work was supported in part by the National Natural Science Foundation of China under Grant 61701245, in part by The Startup Foundation for Introducing Talent of NUIST 2243141701030, in part by A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

- [1] Donoho D. Compressed sensing. *IEEE Transactions on Information Theory* 2006;52(4):1289–306.
- [2] Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med* 2007;58(6):1182–95.
- [3] Liu W, Yin W, Shi L, Duan J, Yu C, Wang D. Under-sampled CS image reconstruction using nonconvex nonsmooth mixed constraints. *Multimed Tools Appl* 2018;1–34.
- [4] Ravishanker S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans Med Imaging* 2012;30(7):964–77.
- [5] Ding X, Paisley J, Huang Y, Chen X, Huang F, Zhang X. Compressed sensing MRI with Bayesian dictionary learning. *IEEE International Conference on Image Processing*. 2013. [Melbourne, Australia].
- [6] Zhan Z, Cai J, Guo D, Liu Y, Chen Z, Qu X. Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. *IEEE Transactions on Biomedical Engineering* 2015;63(9):1850–61.
- [7] Qu X, Guo D, Ning B, Hou Y, Lin Y, Cai S, et al. Undersampled MRI reconstruction with the patch-based directional wavelets. *IEEE Trans Med Imaging* 2014;18(6):843–56.
- [8] Liu Y, Zhan Z, Cai J, Guo D, Chen Z, Qu X. Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging. *IEEE Trans Med Imaging* 2016;35(9):2130–40.
- [9] Huang Y, Paisley J, Lin Q, Ding X, Fu X, Zhang X. Bayesian nonparametric dictionary learning for compressed sensing MRI. *IEEE Trans Image Process* 2014;23(12):5007–19.
- [10] Qu X, Hou Y, Lam F, Guo D, Zhong J, Chen Z. Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator. *Med Image Anal* 2014;18(6):843–56.
- [11] Zhuang P, Zhu X, Ding X. MRI reconstruction with an edge-preserving filtering prior. *Signal Processing* 2019;155:346–57.
- [12] Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, et al. Accelerating magnetic resonance imaging via deep learning. *IEEE International Symposium on Biomedical Imaging* 2016:514–7.
- [13] Schlemper J, Caballero J, Hajnal J, Price A, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging* 2018;37(2):491–503.
- [14] Lee D, Yoo J, Ye J. Deep residual learning for compressed sensing MRI. *IEEE International Symposium on Biomedical Imaging* 2017:15–8.
- [15] Yang Y, Sun J, Li H, Xu Z. ADMM-Net: a deep learning approach for compressive sensing MRI. *arXiv* 2017;1705:6869.
- [16] Schlemper J, Castro C, Wen B, Chen Q, Oktay O, Duan J, et al. Bayesian deep learning for accelerated MR image reconstruction. *International Workshop on Machine Learning for Medical Image Reconstruction*. Cham: Springer; 2018.
- [17] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2016:770–8.
- [18] Lee D, Yoo J, Tak S, Ye J. Deep residual learning for accelerated MRI using magnitude and phase networks. *arXiv* 2018;1804:432.
- [19] Song S, Shim H. Depth reconstruction of translucent objects from a single time-of-flight camera using deep residual networks. *arXiv* 2018;1809:10917.
- [20] Cai C, Zeng Y, Wang C, Cai S, Zhang J, Chen Z, et al. High efficient reconstruction of single-shot T2 mapping from overlapping-echo detachment planar imaging based on deep residual network. *arXiv* 2017;1708:5170.
- [21] Fan Z, Sun L, Ding X, Huang Y, Cai C, Paisley J. A segmentation-aware deep fusion network for compressed sensing MRI. *Proceedings of the European Conference on Computer Vision (ECCV)* 2018:55–70.
- [22] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017;1(2):5.
- [23] Zhang Y, Li K, Li K. Image super-resolution using very deep residual channel attention networks. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
- [24] Kim J, Lee J, Lee K. Accurate image super-resolution using very deep convolutional networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2016:1646–54.
- [25] Jiang D, Dou W, Vosters L, Xu X, Sun Y, Tan T. Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. *Jpn J Radiol* 2018;36(9):566–74.
- [26] Zhang K, Zuo W, Gu S, Zhang L. Learning deep CNN denoiser prior for image restoration. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2017:3929–38.
- [27] Chang L, Shang Z, Qin A. A multiscale image denoising algorithm based on dilated residual convolution network. *arXiv* 2018;1812:9131.
- [28] Fan Z, Wu H, Fu X, Huang Y, Ding X. Residual-guide network for single image deraining. *ACM Multimedia Conference on Multimedia Conference*. 2018. p. 1751–9.
- [29] Fu X, Liang B, Huang Y, Ding X, John P. Lightweight pyramid networks for image deraining. *arXiv* 2018;1805:6173.

- [30] Matsui T, Fujisawa T, Yamaguchi T. Single-image rain removal using residual deep learning. 25th IEEE International Conference on Image Processing (ICIP). 2018. p. 3928–32.
- [31] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. 1511. arXiv; 2015. p. 7122.
- [32] Moeskops P, Pluim J. Isointense infant brain MRI segmentation with a dilated convolutional neural network. 1708. arXiv; 2017. p. 2757.
- [33] Wolterink M, Leiner T, Viergever A, Išgum I. Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease. Reconstruction, segmentation, and analysis of medical images. 2016. p. 95–102.
- [34] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015:1–9.
- [35] Duan J, Bello G, Schlemper J, Bai W, Dawes T, Biff C, et al. Automatic 3D bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. IEEE Trans Med Imaging 2019.
- [36] Bello G, Dawes T, Duan J, Biffi C, Marvaio A, Howard L, et al. Deep-learning cardiac motion analysis for human survival prediction. Nature Machine Intelligence 2019;1(2):95.
- [37] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. 2011. p. 315–23.
- [38] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv 2015. [1502.03167].
- [39] Huang J, Zhang S, Metaxas D. Efficient MR image reconstruction for compressed MR imaging. Med Image Anal 2011:670–9.
- [40] Yang J, Zhang Y, Yin W. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. IEEE Journal of Selected Topics in Signal Processing 2010:288–97.
- [41] Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image video quality assessment. Electron Lett 2008;44(13):800–1.
- [42] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12.
- [43] Knoll F, Hammernik K, Zhang C, Moeller S, Pock T, Sodickson D, et al. Deep learning methods for parallel magnetic resonance image reconstruction. 1904. arXiv; 2019. p. 1112.
- [44] Duan J, Ward WO, Sibbett L, Pan Z, Bai L. Introducing diffusion tensor to high order variational model for image reconstruction. Digital Signal Processing 2017:323–36.
- [45] Lu W, Duan J, Qiu Z, Pan Z, Liu RW, Bai L. Implementation of high-order variational models made easy for image processing. Mathematical Methods in the Applied Sciences 2016;39(14):4208–33.



Yuxiang Dai is a graduate student with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include magnetic resonance imaging and deep learning.



Peixian Zhuang received the Ph.D. degree from Xiamen University, Xiamen, China, in 2016. He is a winner of the Best PhD. Thesis Award in Fujian Province in 2017. He is currently a Lecturer with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include image sparse representation and Bayesian machine learning.