# Multimodal MRI Acceleration via Deep Cascading Networks with Peer-layer-wise Dense Connections

Xiao-Xin Li[1,2], Zhijie Chen[1], Xin-Jie Lou[1], Junwei Yang[3], Yong Chen[4], and Dinggang Shen[2,5,*]

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China
[2] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
[3] Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
[4] Department of Radiology, Case Western Reserve University, Cleveland, Ohio, USA
[5] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

**Abstract.** Medical diagnosis benefits from multimodal Magnetic Resonance Imaging (MRI). However, multimodal MRI has an inherently slow acquisition process. For acceleration, recent studies explored using a fully-sampled side modality (fSM) as a guidance to reconstruct the fully-sampled query modalities (fQMs) from their undersampled $k$-space data via convolutional neural networks. However, even aided by fSM, the reconstruction of fQMs from *highly* undersampled QM data (uQM) is still suffering from aliasing artifacts. To enhance reconstruction quality, we suggest to fully use both uQM and fSM via a deep cascading network, which adopts an iterative Reconstruction-And-Refinement (iRAR) structure. The main limitation of the iRAR structure is that its intermediate reconstruction operators impede the feature flow across subnets and thus leads to *short-term memory*. We therefore propose two typical Peer-layer-wise Dense Connections (PDC), namely, inner PDC (iPDC) and end PDC (ePDC), to achieve *long-term memory*. Extensive experiments on different query modalities under different acceleration rates demonstrate that the deep cascading network equipped with iPDC and ePDC consistently outperforms the state-of-the-art methods and can preserve anatomical structure faithfully up to 12-fold acceleration.

**Keywords:** MRI Acceleration · Guidance-Based Reconstruction Methods · Deep Cascading Networks · Peer-layer-wise Dense Connections

## 1 Introduction

Due to increasing the diversity of diagnostic information in a single examination, multimodal Magnetic Resonance Imaging (MRI) is commonly used in clinical application. However, all of these modalities have to be acquired in a sequential manner and data sampling of each modality in $k$-space is also slow

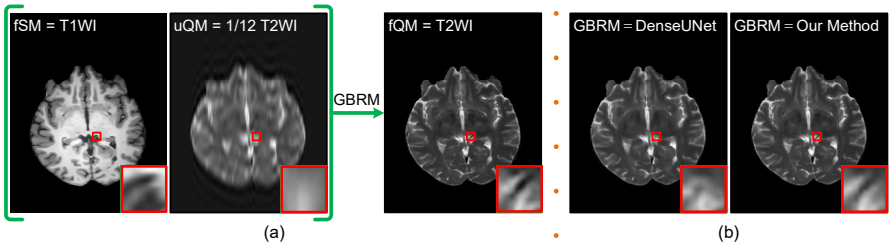


**Fig. 1.** (a) Framework of GBRMs. (b) Predictions of DenseUNet [22] and our method from 1/12 T2WI with the aid of T1WI. Aliasing artifacts can be seen clearly in the prediction of DenseUNet but can be well reduced by our method.

due to MRI physics and physiological constraints. This makes the whole scanning procedure rather lengthy. To accelerate multimodal MRI acquisition, recent studies [20,5,21,11,22,2,4] explored the Guidance-Based Reconstruction Methods (GBRMs), as illustrated in in Fig. 1 (a). That is, a particular modality requiring *shorter* scanning time, e.g., T1-weighted image (T1WI), is first fully sampled and used as a *side modality* (SM) [5,4], and the other modalities requiring *longer* scanning time, e.g., T2-weighted image (T2WI), referred to as *query modalities* (QMs), are then undersampled with a desired acceleration rate and subsequently reconstructed with the aid of the SM data. For simplicity, we denote by fSM, fQM, and uQM the fully-sampled SM data, fully-sampled QM data, and undersampled QM data acquired with the desired acceleration rate, respectively.

However, even aided by fSM and using the state-of-the-art Convolutional Neural Networks (CNNs) [11,21,22,2], the reconstruction result of the existing GBRMs from uQM with a high acceleration rate is still suffering from aliasing artifacts, as shown in the left subfigure of Fig. (b). One main reason is that both fSM and uQM have very low involvement in the forward mapping of the existing CNN models. The existing CNN models [11,21,22,2] were all based on the UNet architecture [15], which induces that fSM and uQM can be only directly used at the very beginning or very end of the network. As such, the sampled $k$-space data in uQM might be distorted heavily with the depth of the network increasing, and the guiding power of fSM will be weakened accordingly.

To fully use fSM and uQM, we suggest to adopt the Deep Cascading Networks [3,16,14], which employ an iterative Reconstruction-And-Refinement (iRAR) structure. In particular, the iRAR structure consists of several cascaded subnets and each subnet performs reconstruction by refining the reconstruction result of its previous subnet. Compared to UNet, the iRAR structure is much more flexible to repeatedly involve both uQM and fSM. Specifically, we can use the acquired $k$-space in uQM by adding a data consistency (DC) layer [16] at the end of each subnet, and reintegrate fSM into the network at the beginning of each subnet. However, the intermediate reconstructions of the iRAR structure impede the features flow from low-level subnets to high-level ones and thus lead to limited receptive field size and *short-term memory*.

To enhance information flow, we introduce two typical *peer-layer-wise dense connections* (PDC), namely inner PDC (iPDC) and end PDC (ePDC), to the iRAR structure. Here, peer layers mean the layers having the same index in different subnets. Image reconstruction benefits from dense connections [18,17,24,23,21,22], as dense connections are helpful for collecting rich knowledge and reducing feature redundancy [9]. For all we know, the existing dense connections are all designed in a *sequential-layer-wise* manner. That is, dense connections are only imposed on the layers/blocks which are cascaded sequentially. Clearly, using sequential-layer-wise dense connections (SDC) is not an efficient way to boost performance for the iRAR structure, as it cannot enlarge receptive field, and can only lead to *restricted long-term memory* [17] (i.e., the low-level features are still confined in the same subnet and cannot flow across subnets). By contrast, PDC can result in long-term memory. Please refer to Fig. 1 of the supplementary material for more clarity about the difference between PDC and SDC. In this work, we introduce iPDC and ePDC to enhance information and gradient flow between the hidden convolutional layers and the outputs of all subnets, respectively.

In summary, the study on the guidance-based multimodal MRI acceleration problem makes our work contribute in the following three aspects. *First*, to fully involve both fSM and uQM, we suggest to use the iRAR structure, instead of the UNet structure, to build GBRMs. *Second*, we create two typical *peer-layer-wise dense connections* (PDC), namely, iPDC and ePDC, to enhance information flow of the iRAR-base network. *Third*, based on iPDC and ePDC, we propose a novel iRAR-based network, namely, a Deep Cascading Network with PDC (DCNwPDC). As demonstrated in Fig. 1 (b) and experiments in Section 4, our DCNwPDC outperforms the state-of-the-art methods for different query modalities under different acceleration rates.

## 2 Problem Formulation

Let $\mathrm{y} \in \mathbb{C}^{M\times N}/\mathrm{y}' \in \mathbb{C}^{M\times N}$ represent fQM/fSM in $k$-space, and $\mathrm{x}/\mathrm{x}'$ be the corresponding 2D MRI image reconstructed from $\mathrm{y}/\mathrm{y}'$ via 2D Inverse Fourier Transform (IFT), $\boldsymbol{F}^{-1}$, i.e., $\mathrm{x} = \boldsymbol{F}^{-1}(\mathrm{y})$ and $\mathrm{x}' = \boldsymbol{F}^{-1}(\mathrm{y}')$. Our problem is to reconstruct x from uQM $\mathrm{y}_u$ with the aid of fSM $\mathrm{x}'/\mathrm{y}'$, such that:

$$\mathrm{y}_u = \mathrm{M} \odot \mathrm{y} = \mathrm{M} \odot \boldsymbol{F}(\mathrm{x}), \tag{1}$$

where $\odot$ is the element-wise multiplication operation, $\boldsymbol{F}$ denotes 2D Fourier Transform (FT), and $\mathrm{M} \in \{0,1\}^{M\times N}$ represents a binary undersampling mask used for MRI acceleration. We define M by using the widely-used horizontal Cartesian undersampling scheme [6,16,22,12], where $k$-space is fully-sampled in the frequency-encoding direction and undersampled in the phase-encoding direction. Fig. 2 in our supplementary materials illustrates the undersampling masks used in this work. As $\mathrm{y}_u$ is highly undersampled, directly applying IFT on $\mathrm{y}_u$ will give rise to a highly aliased reconstruction of x: $\mathrm{x}_u = \boldsymbol{F}^{-1}(\mathrm{y}_u)$. Next, we will explore how to reconstruct x with the aid of $\mathrm{x}'/\mathrm{y}'$ and the proposed DCNwPDC.

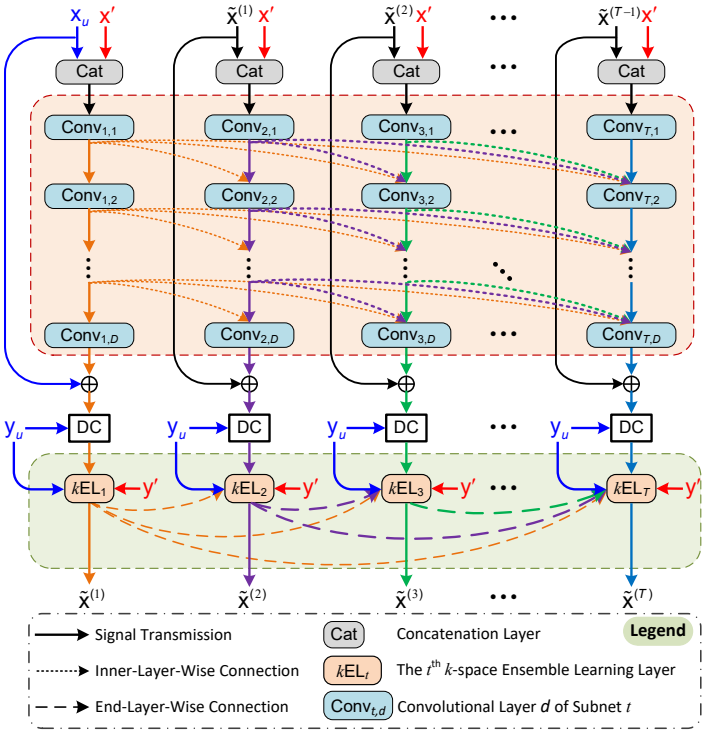


**Fig. 2.** The structure of the proposed Deep Cascading Network with two typical Peer-layer-wise Dense Connections: iPDC (marked by the arrows with dotted lines) and ePDC (marked by the arrows with dashed lines). The red and blue solid arrows indicate how uQM ($\mathrm{x}_u/\mathrm{y}_u$) and fSM ($\mathrm{x}'$) busily participate in the forward mapping.

## 3 Proposed Method

The overall architecture of the proposed DCNwPDC is schematically illustrated in Fig. 2. It consists of $T$ subnets. The whole architecture is well designed to adapt to our guidance-based multimodal MRI acceleration problem and to overcome the problem caused by short-term memory of the iRAR structure. Specially, it comprises the following three elements.

**Long-term Involvement of uQM and fSM via Using the iRAR Structure.** As shown in Fig. 2, our DCNwPDC comprises $T$ subnets: $f_1, f_2, \cdots, f_T$. The long-term involvement of uQM and fSM is implemented by taking $\mathrm{y}_u$, $\mathrm{x}'$ and $\mathrm{y}'$ as input of each subnet $f_t$. Let $\mathrm{x}^{(t)}$ be the output of $f_t$, $\mathrm{x}^{(0)} = \mathrm{x}_u$, and

$[\cdot,\cdot,\cdots]$ denote the concatenation operator. We define $f_t$ in an iterative way:

$$\mathrm{x}_{i\mathrm{CNN}}^{(t)} = \mathrm{x}^{(t-1)} + i\mathrm{CNN}_t\left(\left[\mathrm{x}^{(t-1)}, \mathrm{x}'\right]; \hat{\mathbf{W}}_t\right) \tag{2}$$

$$\mathrm{x}_{\mathrm{DC}}^{(t)} = \mathrm{DC}\left(\mathrm{y}_u, \mathrm{x}_{i\mathrm{CNN}}^{(t)}; \lambda, \mathrm{M}\right) \tag{3}$$

$$\mathrm{x}^{(t)} = k\mathrm{EL}_t\left(\mathrm{y}_u, \mathrm{X}^{(t)}, \mathrm{y}'; \tilde{\mathbf{W}}_t\right), \tag{4}$$

where $\mathrm{X}^{(t)} \triangleq \left\{\mathrm{x}^{(1)}, \cdots, \mathrm{x}^{(t-1)}, \mathrm{x}_{\mathrm{DC}}^{(t)}\right\}$, $\mathrm{x}_{i\mathrm{CNN}}^{(t)}$ and $\mathrm{x}_{\mathrm{DC}}^{(t)}$ are two intermediate outputs of $f_t$, and $\hat{\mathbf{W}}_t/\{\lambda, \mathrm{M}\}/\tilde{\mathbf{W}}_t$ is the parameter set of $i\mathrm{CNN}_t/\mathrm{DC}/k\mathrm{EL}_t$. Specially, $i\mathrm{CNN}_t$ consists of $D$ convolutional layers $\{\mathrm{Conv}_{t,d}\}_{d=1}^{D}$ and goes in image domain; DC denotes the data consistence (DC) layer [16], and its parameter $\lambda$ is decided by the noise level of uQM and set to be zero in this work; $k\mathrm{EL}_t$ is the $k$-space ensemble-learning layer of $f_t$. We next focus on $i\mathrm{CNN}_t$ and $k\mathrm{EL}_t$.

**Long-term Memory of Internal Features via iPDC.** The existing iRAR-based networks only have short-term memory [3,16] or restricted long-term memory [14,13]. To improve the information flow between subnets, we propose a different connectivity pattern: inner Peer-layer-wise Dense Connections (iPDC). Here, peer layers mean the layers having the same index in different subnets. For instance, $\{\mathrm{Conv}_{\iota,d}\}_{\iota=1}^{T}$ forms the peer-layer set of $\mathrm{Conv}_{t,d}$ in Fig. 2. For any inner layer $\mathrm{Conv}_{t,d}$ ($d \in \{2, 3, \cdots, D-1\}$), iPDC feeds its output forward to its child peer-layer set $\{\mathrm{Conv}_{t+\iota,d+1}\}_{\iota=0}^{\tau}$. Here, $\tau = \min(T-t, \delta-1)$, and $\delta \leq T$ is the *maximum memory length* of iPDC, which is predefined as the maximum number of the forward connections between a layer and its child peer-layers. Consequently, $\mathrm{Conv}_{t,d+1}$ receives the feature-maps from its parent peer-layer set $\{\mathrm{Conv}_{t+\iota,d}\}_{\iota=-\ell}^{0}$ ($\ell = \min(t, \delta-1)$) as input:

$$\begin{aligned}\mathrm{H}_{t,d+1} &= \mathrm{Conv}_{t,d+1}\left(\mathrm{H}_{t-\ell,d}, \cdots, \mathrm{H}_{t-1,d}, \mathrm{H}_{t,d}; \hat{\mathbf{W}}_{t,d+1}\right) \\ &= \hat{\mathrm{W}}_{t,d+1}^{t} * \left[\hat{\mathrm{W}}_{t,d+1}^{t-\ell} * \mathrm{H}_{t-\ell,d}, \cdots, \hat{\mathrm{W}}_{t,d+1}^{t-1} * \mathrm{H}_{t-1,d}, \sigma\left(\mathrm{H}_{t,d}\right)\right],\end{aligned} \tag{5}$$

where $\mathrm{H}_{t,d}$ is the output feature map of $\mathrm{Conv}_{t,d}$, $*$ represents convolutional operator, $\sigma$ denotes the rectifier linear unit (ReLU), and $\hat{\mathbf{W}}_{t,d+1} \triangleq \left\{\hat{\mathrm{W}}_{t,d+1}^{\iota}\right\}_{\iota=1}^{t}$. There are two points worth noting in Eq. (5). First, ReLU is only applied to $\mathrm{H}_{t,d}$. That is, the feature-maps produced by the parent peer layers of $\mathrm{Conv}_{t,d+1}$, except for $\mathrm{Conv}_{t,d}$, should not be activated in advance. We experimentally find that early activation by ReLU might lead to performance degradation. Second, the connection way of PDC is different from that of DenseNet [9], where the feature-maps of preceding layers are directly concatenated and fed to the target layer. PDC first filters input feature-maps by convolution, which adaptively controls how much of the previous memories should be reserved.

**Long-term Memory of Intermediate Predictions via ePDC and *k*-space Ensemble Learning**. Recent research [10,17,19] demonstrated that Ensemble Learning (EL) can significantly boost the reconstruction performance.

Also, $k$-space learning [6,7] is a useful supplement to boost MRI reconstruction. Inspired by these works, we use a $k$-space Ensemble Learning ($k$EL) layer at the end of each subnet $f_t$ to fully integrate the input predictions,

$$\begin{aligned}\tilde{\mathrm{x}}^{(t)} &= k\mathrm{EL}_t\left(\mathrm{y}_u, \mathrm{X}^{(t)}, \mathrm{y}'; \tilde{\mathbf{W}}_t\right), \\ &= \mathrm{DC}\left(\mathrm{y}_u, k\mathrm{CNN}_t\left(\left[\mathrm{WS}\left(\mathrm{X}^{(t)}; \mathrm{w}_t\right), \mathrm{y}'\right]; \tilde{\mathbf{W}}_t^k\right); \lambda, \mathrm{M}\right). \end{aligned} \tag{6}$$

Here $\tilde{\mathbf{W}}_t = \mathrm{w}_t \cup \tilde{\mathbf{W}}_t^k$, $\mathrm{w}_t/\tilde{\mathbf{W}}_t^k$ is the parameter set of $\mathrm{WS}(\cdot)/k\mathrm{CNN}_t$. Specially, $\mathrm{WS}(\cdot)$ calculates the weighted sum of the elements in $\mathrm{X}^{(t)}$, and $k\mathrm{CNN}_t$ consists of three convolutional layers, which have the same configurations with $\mathrm{Conv}_{t,1}$, $\mathrm{Conv}_{t,2}$, and $\mathrm{Conv}_{t,D}$, and performs in $k$-space with the aid of $\mathrm{y}'$. Note that $k\mathrm{EL}_t$ is beyond the simple weighted sum of the input predictions, which might lead to unstable performance. The result of the weighted sum is further processed by $k\mathrm{CNN}_t$, which also provides a chance to involve $\mathrm{y}'$.

**Network Training.** Given a training dataset $\left\{\mathrm{x}_u^{\langle\gamma\rangle}, \mathrm{y}_u^{\langle\gamma\rangle}, \mathrm{x}'^{\langle\gamma\rangle}, \mathrm{y}'^{\langle\gamma\rangle}, \mathrm{x}^{\langle\gamma\rangle}\right\}_{\gamma=1}^{\Gamma}$, our goal is to find the best model $f_{\mathrm{DCNwPDC}}$ that accurately reconstructs x from $(\mathrm{x}_u, \mathrm{y}_u, \mathrm{x}', \mathrm{y}')$. Note that $f_{\mathrm{DCNwPDC}}$ outputs multiple intermediate predictions. To boost performance, existing work [10,17] usually adopt the multi-supervision strategy. However, multi-supervision will introduce additional weight parameters to balance all predictions, and these weight parameters should be carefully chosen. For our network, multi-supervision is unnecessary, as all intermediate predictions are connected to the final output layer, which leads to implicit multi-supervision [9]. We therefore directly minimize $\frac{1}{2}\left\|\mathrm{x} - f_{\mathrm{DCNwPDC}}\left(\mathrm{x}_u, \mathrm{y}_u, \mathrm{x}', \mathrm{y}'\right)\right\|_2^2$ averaged over the whole training set. We train our model using the Adam solver with a momentum of 0.9 and perform 100 epochs. The batch size is set to 6 and the initial learning rate is set to $10^{-4}$, which is divided by 10 for every 50 epochs. The network is implemented in Python using PyTorch library. Our source code will be subsequently published online.

## 4 Experiments

We use the MSSEG dataset [1] to evaluate the efficacy of our DCNwPDC. MSSEG includes four MRI modalities: T1WI, T2WI, DP and FLAIR, which are scanned from three different scanners. We use T1WI as the side modality, and the rest T2WI, DP and FLAIR as the query modalities. As the three scanners generate images of different sizes, we select five subjects scanned by the same Philips Ingenia 3T scanner, which produces 3D volumes containing 336 $\times$ 336 $\times$ 261 voxels. By excluding the slices only containing the background, we obtain around 200 samples in axial direction for each subject. Training and test subjects are split to 4 : 1. We evaluate the reconstruction results using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Throughout the experiments, we use D5C5 [16] as a baseline network to implement our DCNwPDC. D5C5 is a classical iRAR-based network. By setting $D = 5$ and $T = 5$, we have DCNwPDC = D5C5 + iPDC + $k$EL.

**Table 1.** Effect of iPDC and $k$EL for reconstruction of 1/8 T2WI assisted by T1WI. $\sim$ and $\sim\sim$ denote D5C5 and D5C5(Conv+Cat, 3), respectively.

| | Subject 01 | | Subject 02 | | Subject 03 | | Average | |
|---|---|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| $\sim$ | 37.54 | 0.9895 | 38.94 | 0.9922 | 38.67 | 0.9920 | 38.38 | 0.9912 |
| $\sim$+iSDC(Cat, 4) | 37.82 | 0.9903 | 38.74 | 0.9920 | 38.96 | 0.9925 | 38.51 | 0.9916 |
| $\sim$+iPDC(Cat, 2) | 37.67 | 0.9898 | 39.02 | 0.9925 | 38.93 | 0.9924 | 38.54 | 0.9916 |
| $\sim$+iPDC(Res, 2) | 37.56 | 0.9897 | 38.78 | 0.9921 | 38.74 | 0.9921 | 38.36 | 0.9913 |
| $\sim$+iPDC(Conv+Cat, 2) | 37.76 | 0.9899 | 39.00 | 0.9925 | 39.06 | 0.9927 | 38.61 | 0.9917 |
| $\sim$+iPDC(ReLU+Conv+Cat, 2) | 37.72 | 0.9900 | 38.91 | 0.9925 | 38.95 | 0.9926 | 38.53 | 0.9917 |
| $\sim$+iPDC(Conv+Cat, 3) | 37.84 | 0.9903 | 39.10 | 0.9926 | 38.95 | 0.9926 | 38.63 | 0.9918 |
| $\sim$+iPDC(Conv+Cat, 4) | 37.84 | **0.9904** | **39.12** | **0.9928** | **39.10** | **0.9929** | **38.69** | **0.9920** |
| $\sim$+iPDC(Conv+Cat, 5) | **37.86** | **0.9904** | **39.12** | **0.9928** | 39.09 | 0.9928 | **38.69** | **0.9920** |
| $\sim\sim$+WS(X) | 37.80 | 0.9903 | 39.12 | 0.9929 | 39.13 | 0.9929 | 38.68 | 0.9920 |
| $\sim\sim+k$CNN($x_{DC}$) | 37.90 | 0.9905 | 39.22 | 0.9930 | 39.12 | 0.9929 | 38.75 | 0.9921 |
| $\sim\sim+k$CNN($x_{DC}$, y′) | 37.87 | 0.9904 | 39.24 | 0.9930 | 39.22 | 0.9930 | 38.78 | 0.9921 |
| $\sim\sim+k$CNN(WS(X)) | 37.84 | 0.9904 | 39.23 | 0.9930 | 39.13 | 0.9929 | 38.73 | 0.9921 |
| $\sim\sim+k$CNN(WS(X), y′) | **37.95** | **0.9906** | **39.26** | **0.9931** | **39.26** | **0.9933** | **38.82** | **0.9923** |

We first evaluate the effects of two important elements of $f_{\text{DCNwPDC}}$: iPDC and $k$EL. We fix T1WI and 1/8 undersampled T2WI (shortly, 1/8 T2WI) as fSM and uQM, and sequentially select Subjects 01, 02 and 03 as the test subject.

**Effect of iPDC.** We compare iPDC defined in Eq. (5) against iSDC (inner SDC) and the variants of iPDC with other possible configurations. Eq. (5) indicates that iPDC mainly has two configurations: the connection way between the input feature-maps and the maximum memory length $\delta$. Eq. (5) uses the Conv+Cat connections. The other possible operators that can replace Conv+Cat are are Cat, Res (Residual) [8], and ReLU+Conv+Cat. We represent a specific iPDC/iSDC by iPDC/iSDC(Connection-way, $\delta$). The top sub-table of Table 1 reports how the connection way and $\delta$ influence the reconstruction performance. We can find that compared to $\sim$ ($\sim$ denotes D5C5), both $\sim$+iSDC(Cat, 4) and $\sim$+iPDC(Res, 2) lead to performance degradation on Subject 02. By comparing $\sim$+iPDC(Conv+Cat, 2) and $\sim$+iPDC(ReLU+Conv+Cat, 2), we can find early ReLU might lead to performance dropping. $\delta$ usually boosts performance but a too big $\delta$ might lead to overfitting due to using more parameters. Table 1 also indicates our iPDC can effectively use the network capacity by comparing the performance of $\sim$, $\sim$+iSDC(Cat, 4) and $\sim$+iPDC(Cat, 2). Clearly, $\sim$+iSDC(Cat, 4) uses more parameters than $\sim$ but it might lead to performance degeneration against $\sim$. By contrast, $\sim$+iPDC(Cat, 2) leads to stable performance boost over $\sim$, although it has fewer parameters than $\sim$+iSDC(Cat, 4). This demonstrates the superiority of the network design of our iPDC against iSDC.

**Effect of $k$EL.** By observing Eq. (6), we can obtain several options to implement the main part (except DC) of $k$EL: e.g., WS(X), $k$CNN($\text{x}_{\text{DC}}$), $k$CNN($\text{x}_{\text{DC}}, \text{y}'$) and etc. The bottom sub-table of Table 1 reports the related performance. Only $k\text{CNN}\left(\text{WS(X)}, \text{y}'\right)$, as defined in Eq. (6), stably outperforms the other options.

**Comparisons with Existing Methods.** We compare our DCNwPDC with the basic zero-filling method and four advanced CNN-based methods: UNet [15],

**Table 2.** Quantitative evaluation of the compared methods in Regions of Interest (ROI) for the reconstruction performance of 1/4, 1/8, and 1/12 undersampled QM $\in$ {T2WI, DP} assisted by T1WI. Five-fold cross validations are performed. Numbers in **red** and blue indicate the best and second best performance, respectively.

| | 1/4 T2WI | | 1/8 T2WI | | 1/12 T2WI | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Zero Filling | 31.01±0.0671 | 0.9591±4.35e-06 | 29.47±0.61603 | 0.9038±3.94e-05 | 27.19±1.2760 | 0.8683±4.90e-05 |
| UNet | 37.48±0.0350 | 0.9784±8.27e-08 | 36.72±0.41053 | 0.9646±9.95e-07 | 35.69±0.6826 | 0.9632±1.19e-06 |
| DenseUNet | 37.83±0.0761 | 0.9785±2.45e-07 | 36.82±0.40868 | 0.9645±1.22e-06 | 35.76±0.6526 | 0.9631±1.15e-06 |
| RefineGAN | 37.93±0.0405 | 0.9776±0.83e-07 | 36.94±0.46610 | 0.9637±1.12e-06 | 36.16±0.5782 | 0.9624±1.21e-06 |
| D5C5 | 39.04±0.0242 | 0.9785±4.57e-07 | 37.71±0.38307 | 0.9652±1.16e-06 | 36.73±0.5317 | 0.9637±1.26e-06 |
| D5C5+iPDC | 39.53±0.0420 | 0.9792±7.95e-08 | 38.25±0.38315 | 0.9663±9.03e-07 | 37.38±0.4986 | 0.9645±9.25e-07 |
| D5C5+iPDC+$k$EL | **39.79±0.0338** | **0.9795±6.83e-08** | **38.39±0.33928** | **0.9668±8.13e-07** | **37.45±0.4882** | **0.9648±9.17e-07** |
| | 1/4 DP | | 1/8 DP | | 1/12 DP | |
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Zero Filling | 33.17±0.0683 | 0.9692±3.98e-06 | 32.10±0.64329 | 0.9417±3.25e-05 | 29.21±0.6247 | 0.9281±5.42e-05 |
| UNet | 38.26±0.0889 | 0.9779±1.69e-07 | 37.72±0.44730 | 0.9655±1.29e-06 | 36.35±0.5251 | 0.9613±1.73e-06 |
| DenseUNet | 38.38±0.0791 | 0.9777±1.39e-07 | 38.33±0.45025 | 0.9653±1.12e-06 | 36.52±0.9043 | 0.9613±1.96e-06 |
| RefineGAN | 39.04±0.0354 | 0.9777±1.97e-07 | 39.93±0.49513 | 0.9668±1.63e-06 | 37.85±0.2215 | 0.9638±1.92e-06 |
| D5C5 | 42.48±0.0586 | 0.9795±1.35e-07 | 40.92±0.35283 | 0.9680±1.47e-06 | 38.78±0.7644 | 0.9654±1.64e-06 |
| D5C5+iPDC | 42.83±0.0726 | 0.9803±8.59e-08 | 41.40±0.32191 | 0.9686±9.12e-07 | 39.27±0.9208 | 0.9659±7.17e-07 |
| D5C5+iPDC+$k$EL | **42.87±0.0540** | **0.9806±8.31e-08** | **41.51±0.30215** | **0.9690±8.49e-07** | **39.41±0.9027** | **0.9661±7.69e-07** |

DenseUNet [22], RefineGAN [14], and D5C5 [16]. Note that even if the GAN-based networks [11,2] are specifically designed for guided MRI reconstruction, we choose RefineGAN [14] for comparison due to its stronger reconstruction abilities. Specially, RefineGAN also adopts the iRAR structure by cascading two UNets. For fair comparison, we implement all iRAR-based methods by inputting fSM at the beginning of all subnets. Table 2 reports the quantitative evaluation results of the compared methods on two query modalities, T2WI and DP, in Regions of Interest (ROI). Note that we perform quantitative evaluations in ROI rather than in the whole image, as radiologists mainly concern the subtle abnormalities in ROI. Fig. 3 in our supplementary material plots the ROI used for the query modalities. For each uQM acquired with three acceleration rates, our method (i.e., D5C5 + iPDC + $k$EL) and its simplified variant D5C5 + iPDC achieve the best and second best performance, respectively. In particular, our method gains better performance for higher acceleration rates. It indicates that enhancing information flow is particularly important for highly aggressive undersampling. Fig. 3 further demonstrates the superiority of our method by visually comparing the reconstruction quality of the compared methods for 12-fold undersampled T2WI and DP. At this high acceleration rate, we can see clearly that our method better restores more anatomical details than the other compared methods. The quantity and quality evaluation results of the compared methods on FLAIR reconstruction are reported in our supplementary materials.

## 5 Conclusion

By integrating two typical peer-layer-wise dense connections (iPDC and ePDC) to the iRAR-based structure, we propose a novel deep cascading network, which

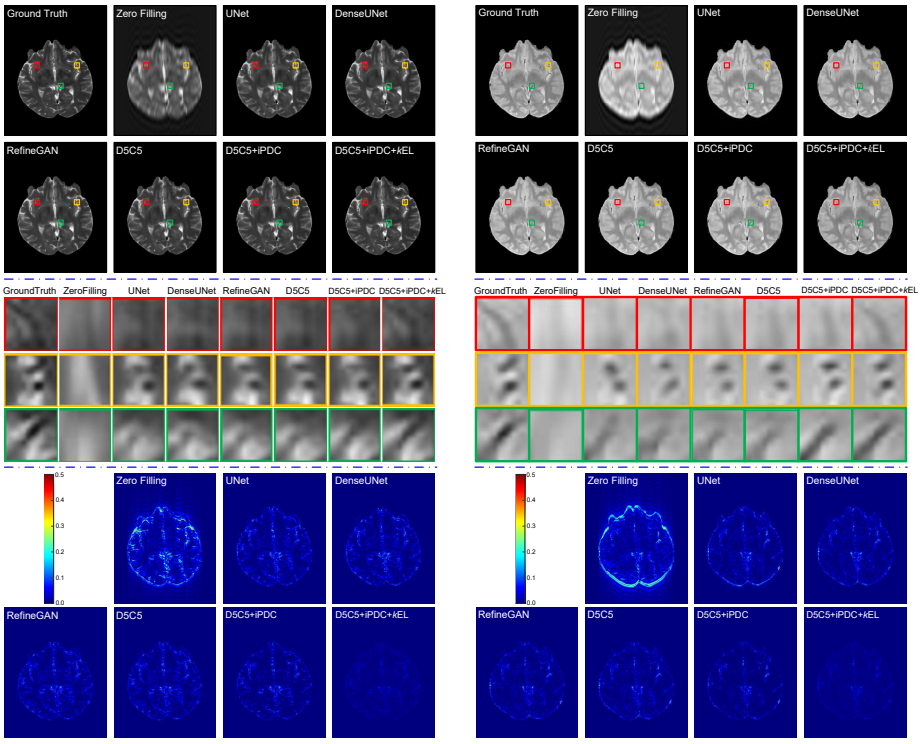


**Fig. 3.** Visual comparison of the reconstruction results (top row) and the error maps (bottom row) of the seven compared methods for 12-fold undersampled T2WI (left) and DP (right). Three subregions of interest for each reconstructed image are enlarged in the middle row for more clarity.

outperforms the state-of-the-art methods. With the proposed network, the acquisition of query modalities can be accelerated up to 12 folds without evident sacrifice of image quality. Future studies will be focused on further improvement of DCNwPDC by optimizing the design of PDC, and extending to handle more sampling patterns beyond the Cartesian trajectory, such as radial, spiral, etc.

## References

1. Commowick, O., Cervenansky, F., Ameli, R.: MSSEG challenge proceedings: multiple sclerosis lesions segmentation challenge using a data management and processing infrastructure. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2016)
2. Dar, S.U., Yurt, M., Shahdloo, M., Ildız, M.E., Tınaz, B., Çukur, T.: Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks. IEEE Journal of Selected Topics in Signal Processing **14**(6), 1072–1087 (2020)

3. Denton, E.L., Chintala, S., Fergus, R., et al.: Deep generative image models using a Laplacian pyramid of adversarial networks. In: Advances in neural information processing systems. pp. 1486–1494 (2015)
4. Ehrhardt, M.J.: Multi-modality imaging with structure-promoting regularisers. arXiv preprint arXiv:2007.11689 (2020)
5. Ehrhardt, M.J., Betcke, M.M.: Multicontrast MRI reconstruction with structure-guided total variation. SIAM Journal on Imaging Sciences **9**(3), 1084–1106 (2016)
6. Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.J., Hwang, D.: KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magnetic resonance in medicine **80**(5), 2188–2201 (2018)
7. Han, Y., Sunwoo, L., Ye, J.C.: k-space deep learning for accelerated MRI. IEEE Trans. Medical Imaging **39**(2), 377–386 (2019)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778. Las Vegas, USA (2016)
9. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708 (2017)
10. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1637–1645 (2016)
11. Kim, K.H., Do, W.J., Park, S.H.: Improving resolution of mr images with an adversarial network incorporating images with different contrast. Medical physics **45**(7), 3120–3131 (2018)
12. Li, W., Feng, X., An, H., Ng, X.Y., Zhang, Y.J.: MRI reconstruction with interpretable pixel-wise operations using reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 792–799 (2020)
13. Qin, C., Schlemper, J., Caballero, J., Price, A.N., Hajnal, J.V., Rueckert, D.: Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE Transactions on Medical Imaging **38**(1), 280–290 (2019)
14. Quan, T.M., Nguyen-Duc, T., Jeong, W.K.: Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. IEEE transactions on medical imaging **37**(6), 1488–1497 (2018)
15. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
16. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic mr image reconstruction. IEEE transactions on Medical Imaging **37**(2), 491–503 (2018)
17. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: IEEE International Conference on Computer Vision. pp. 4539–4547. IEEE International Conference on Computer Vision (2017)
18. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: The IEEE International Conference on Computer Vision (2017)
19. Wang, L., Huang, Z., Gong, Y., Pan, C.: Ensemble based deep networks for image super-resolution. Pattern Recognition **68**, 191–198 (2017)
20. Weizman, L., Eldar, Y.C., Ben Bashat, D.: Reference-based MRI. Medical physics **43**(10), 5357–5369 (2016)
21. Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., Shen, D.: Ultra-fast T2-weighted MR reconstruction using complementary T1-weighted information.

In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 215–223 (2018)
22. Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., Shen, D.: Deep-learning-based multi-modal fusion for fast mr reconstruction. IEEE Transactions on Biomedical Engineering **66**(7), 2105–2114 (2019)
23. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence p. online (2020)
24. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2472–2481 (2018)