Deep Unfolding Convolutional Dictionary Model for Multi-Contrast MRI Super-resolution and Reconstruction

Pengcheng Lei\textsuperscript{1}, Faming Fang\textsuperscript{1}, Guixu Zhang\textsuperscript{1*} and Ming Xu\textsuperscript{2*}

\textsuperscript{1}School of Computer Science and Technology, East China Normal University
\textsuperscript{2}Software Engineering Institute, East China Normal University
pengchenglei1995@163.com, \{fmfang, gxzhang, mxu\}@cs.ecnu.edu.cn

Abstract

Magnetic resonance imaging (MRI) tasks often involve multiple contrasts. Recently, numerous deep learning-based multi-contrast MRI super-resolution (SR) and reconstruction methods have been proposed to explore the complementary information from the multi-contrast images. However, these methods either construct parameter-sharing networks or manually design fusion rules, failing to accurately model the correlations between multi-contrast images and lacking certain interpretations. In this paper, we propose a multi-contrast convolutional dictionary (MC-CDic) model under the guidance of the optimization algorithm with a well-designed data fidelity term. Specifically, we build an observation model for the multi-contrast MR images to explicitly model the multi-contrast images as common features and unique features. In this way, only the useful information in the reference image can be transferred to the target image, while the inconsistent information will be ignored. We employ the proximal gradient algorithm to optimize the model and unroll the iterative steps into a deep CDic model. Especially, the proximal operators are replaced by learnable ResNet. In addition, multi-scale dictionaries are introduced to further improve the model performance. We test our MC-CDic model on multi-contrast MRI SR and reconstruction tasks. Experimental results demonstrate the superior performance of the proposed MC-CDic model against existing SOTA methods. Code is available at https://github.com/lpcccc-cv/MC-CDic.

1 Introduction

Magnetic Resonance Imaging (MRI) is a noninvasive and non-ionizing medical imaging technique widely used for medical diagnosis, clinical analysis, and staging of disease [Song \textit{et al.}, 2019a]. However, due to the physics and physiological constraints, the MRI scanning procedure is always time-consuming. The scanning time of each patient can take up to ten minutes long, which seriously affects the patient experience and leads to high costs [Feng \textit{et al.}, 2022]. To accelerate the MRI procedure, a commonly-used approach is to undersample the k-space measurements and then use the reconstruction algorithm to restore its fully-sampled imaging. The recent mainstream methods for accelerating MRI include MRI reconstruction and MRI super-resolution (SR). MRI reconstruction aims to remove the aliasing artifacts caused by k-space undersampling. MRI SR aims to reconstruct high-resolution (HR) MR images from low-resolution (LR) images. Both two methods can effectively improve the quality of MR images and accelerate MRI without upgrading hardware facilities.

In the past few years, numerous MRI SR [Zhao \textit{et al.}, 2019; Wang \textit{et al.}, 2021; Zhang \textit{et al.}, 2021b] and reconstruction [Wang \textit{et al.}, 2016; Zheng \textit{et al.}, 2019; Nitski \textit{et al.}, 2020] methods have been proposed. However, they only use single-contrast MR images, ignoring complementary information from multi-contrast images. In clinical practice, scanners usually produce exquisite images of body parts with different contrasts, such as T1-weighted, T2-weighted, and proton density weighted images (PD). These contrasts capture the information with the same underlying anatomy and thus exhibit similar structures [Lyu \textit{et al.}, 2020]. Furthermore, since the repetition time and echo time of different contrasts are different. For example, the scanning time of T1 and PD are significantly shorter than those of T2, so it is promising to leverage an HR (or fully sampled) reference image with a shorter acquisition time to reconstruct the modality with a
longer scanning time from an LR (or undersampled) image. We name them guided-SR and guided-reconstruction.

Recently, deep convolutional neural network (DCNN) [Liu and Lei, 2021] has become a mainstream approach for solving MRI guided-SR problems. [Lyu et al., 2020] proposed to fuse the multi-contrast information in the high-level feature space. [Feng et al., 2021] used a multi-stage integration network to explore the rich relations among different contrasts. [Li et al., 2022; Fang et al., 2022] designed a transformer-based framework to fully explore the long-range dependencies in both reference and target images. For solving the multi-contrast reconstruction problem, [Xiang et al., 2018] fused multi-contrast MR acquisitions through deep networks to accelerate MR imaging. [Zhang, 2020] employed an auxiliary contrast as prior information to accelerate MR imaging. [Feng et al., 2022] proposed a multi-modal transformer (MTrans) to transfer multiscale features from the target modality to the auxiliary one. Although these methods have got promising results, their network architectures are black-boxes, making it difficult to understand the underlying mechanism and lack certain physical interpretations. As is well known, the model-based method [Song et al., 2019b] is explainable and easy to incorporate domain knowledge. Most recently, numerous well-designed model-driven DCNN methods have been proposed to solve various image restoration problems, including deblurring [Wang et al., 2020], image fusion [Liu et al., 2019], and so on. For MR images, [Yang et al., 2018; Aggarwal et al., 2018; Sriram et al., 2020] proposed model-based reconstruction methods based on the compressed sensing theory. However, they are designed for single-contrast MRI reconstruction.

In this paper, we propose a novel multi-contrast convolutional dictionary (MC-CDic) model to explicitly model the correlations of multi-contrast images for guided-SR and reconstruction tasks. MC-CDic model is designed by the following two observations. First, the multi-contrast images are sampled under the same anatomical structure, they thus share some common features, e.g., edges and texture information. Second, since the two images are generated by different scan settings, they also have their unique contrast information and some unaligned inconsistent structures, as shown in Figure 1. To better reconstruct the target image, we hope to transfer the common texture information from the reference image to the target image while avoiding the interference of inconsistent information. Taking these into consideration, we first extract the common features and unique features from the multi-contrast images and then fuse these features to generate the desired high-quality images.

To extract the features from both images, firstly, we adopt the convolutional sparse coding (CSC) technique to model the multi-contrast MR images. Specifically, the multi-contrast images are decomposed into two types of features (i.e., common features and unique features) using two types of dictionaries (i.e., common filters and unique filters). Secondly, we employ implicit regularizations on the decomposed features and use the proximal gradient algorithm to optimize them. Thirdly, we unfold the iterative optimization process into a deep network. Especially, the unfolding process strictly follows the rules of the solution and the proximal operators are replaced by learnable residual blocks. Thus all the parameters can be learned in an end-to-end manner. Fourthly, we further employ multi-scale dictionaries to extract the multi-scale sparse representations in each iteration. In this way, the proposed network is not only interpretable but also powerful. Our contributions can be summarized as follows.

- We propose a deep unfolding network based on convolutional dictionary learning for multi-contrast MRI SR and reconstruction. Different from existing methods that design unsophisticated fusion rules, we develop priors and constraints plugged into the model according to the characteristics of multi-contrast MR images.
- We build an observation model for multi-contrast MRI using the CSC technique and use the proximal gradient algorithm to solve the model. Based on the iterative solutions, we unroll the iteration steps into a multi-scale convolutional dictionary network and the proximal operators are replaced by learnable deep residual networks.
- The proposed model combines the good interpretability of the dictionary model with powerful deep neural networks. We test it on guided MRI SR and reconstruction tasks. Extensive experiments on the benchmark datasets demonstrate the effectiveness of the proposed model.

2 MC-CDic Model for MR Images

2.1 Model Formulation

Let \( \hat{x}_1 \in \mathbb{R}^{m \times n} \) and \( \hat{x}_2 \in \mathbb{R}^{m \times n} \) denote two classes of HR (or fully-sampled) MR images with different contrasts (e.g. PD and T2). The corresponding LR (or undersampled) images of \( \hat{x}_2 \) is \( x_2 \). For the guided-SR task, our MC-CDic model aims to restore the HR image \( x_2 \) from LR image \( x_2 \) with the aid of the reference HR image \( \hat{x}_1 \). Since the two contrast images are sampled under the same anatomical structure, they share some common features. Since they are sampled under different scan setting, they also have their unique features. Based on these observations, inspired by [Deng and Dragotti, 2020], we model the multi-contrast images as:

\[
\hat{x}_1 = \sum_k d_k^c \otimes c_k + \sum_k d_k^u \otimes u_k, \tag{1}
\]

\[
x_2 = \sum_k h_k^c \otimes c_k + \sum_k h_k^u \otimes v_k, \tag{2}
\]

where \( \otimes \) denotes the convolutional operation. \( \{d_k^c\}_{k=1}^K \) and \( \{h_k^c\}_{k=1}^K \) are the common filters of \( \hat{x}_1 \) and \( x_2 \), \( \{d_k^u\}_{k=1}^K \) and \( \{h_k^u\}_{k=1}^K \) are the unique filters of \( \hat{x}_1 \) and \( x_2 \). \( \{c_k\}_{k=1}^K \) and \( \{v\}_{k=1}^K \) represent the common features between \( \hat{x}_1 \) and \( x_2 \). \( K \) denotes the channel numbers of the feature representations. Based on the decomposed common features and unique features, the
reconstructed HR image can be represented as:

$$x'_2 = \sum_k q_k \otimes c_k + \sum_k q_k \otimes v_k,$$

(3)

where \{q_k\}_{k=1}^K \text{ and } \{q_k\}_{k=1}^K \text{ are the common filters and the unique filters for reconstructing the HR images } x'_2.

For the above modeling, the main problem is to extract the common features and the unique features from the input images. Assuming all the filters are known, they can be solved by optimizing the following problem:

$$\min_{(C,U,V)} \frac{1}{2} \|x_1 - \sum_k (d_k \otimes c_k + d_k \otimes u_k)\|_F^2 + \lambda_C \psi_C(U) + \lambda_U \psi_U(U),$$

(4)

$$+ \lambda_V \psi_V(V),$$

where \(\{C,U,V\} \in \mathbb{R}^{n \times n \times K}\) are the stacked features response to \(\{c,u,v\}_{k=1}^K\), \(\lambda_C\), \(\lambda_U\), \(\lambda_V\) are the trade-off parameters, \(\psi_C(\cdot), \psi_U(\cdot), \psi_V(\cdot)\) are the regularizers to deliver the prior information to the feature maps.

### 2.2 Optimization Algorithm

To solve this problem, we alternately update each variable with other variables fixed. It leads to the following three subproblems:

\[
\min_{U} \frac{1}{2} \|x_1 - \sum_k (d_k \otimes c_k + d_k \otimes u_k)\|_F^2 + \lambda_U \psi_U(U), \tag{5}
\]

\[
\min_{C} \frac{1}{2} \|x_2 - \sum_k (h_k \otimes c_k + h_k \otimes v_k)\|_F^2 + \lambda_C \psi_C(C), \tag{6}
\]

\[
\min_{V} \frac{1}{2} \|x_1 - \sum_k (d_k \otimes c_k + d_k \otimes u_k)\|_F^2 + \lambda_V \psi_V(V), \tag{7}
\]

Next, we optimize these three subproblems separately using the proximal gradient method. The details are as follows:

**Updating U.** The unique feature of \(x_1\), \(U\) can be updated by solving the quadratic approximation [Beck and Teboulle, 2009] of the subproblem (5) as:

$$\min_U \frac{1}{2} \|U - (U^{t-1} - \eta_u \nabla f(U^{t-1}))\|_F^2 + \lambda_U \psi_U(U),$$

(8)

where \(U^{t-1}\) is the updating result of the last iteration, \(\eta_u\) is the stepsize parameter, and

$$f(U^{t-1}) = \frac{1}{2} \|x_1 - \sum_k (d_k \otimes c_k + d_k \otimes u_k^{t-1})\|_F^2,$$

(9)

$$\nabla f(U^{t-1}) = D^u \otimes \left( \sum_k (d_k \otimes c_k + d_k \otimes u_k^{t-1} - x_1) \right),$$

(10)

where \(\otimes^T\) denotes the transposed convolution, \(D^u \in \mathbb{R}^{1 \times n \times n \times K}\) is a 4-D tensor stacked by \(\{d_k\}_{k=1}^K\). Eq. (10) can be re-writed as:

$$\nabla f(U^{t-1}) = D^u \otimes \left( D^c \otimes C^{t-1} + D^u \otimes U^{t-1} \right) - D^u \otimes x_1,$$

(11)

where \(D^c \in \mathbb{R}^{1 \times n \times n \times K}\) is a 4-D tensor stacked by \(\{d_k^c\}_{k=1}^K\). Corresponding to general regularization term [Donoho, 1995], the solution of Eq. (8) is:

$$U^t = \text{prox}_{\eta_u \lambda_u}(U^{t-1} - \eta_u \nabla f(U^{t-1})), $$

(12)

where \(\text{prox}_{\eta_u \lambda_u}(\cdot)\) is the proximal operator dependent on the regularization term \(\psi_u(\cdot)\) with respect to \(U\). Instead of choosing a fixed regularizer in the model, the form of the proximal operator can be automatically learned from training data.

**Updating V.** Similarly, the quadratic approximation of the problem (6) with respect to \(V\) is:

$$\min_V \frac{1}{2} \|V - (V^{t-1} - \eta_v \nabla f(V^{t-1}))\|_F^2 + \lambda_V \psi_V(V),$$

(13)

where

$$f(V^{t-1}) = \frac{1}{2} \|x_2 - \sum_k (h_k \otimes c_k + h_k \otimes v_k)\|_F^2,$$

(14)

$$\nabla f(V^{t-1}) = H^v \otimes \left( \sum_k (h_k \otimes c_k + h_k \otimes v_k^{t-1}) - x_2 \right).$$

(15)

\(H^v \in \mathbb{R}^{1 \times n \times n \times K}\) is a 4-D tensor stacked by \(\{h_k\}_{k=1}^K\). Eq. 15 can be re-writed as:

$$\nabla f(V^{t-1}) = H^v \otimes \left( H^c \otimes C^{t-1} + H^u \otimes V^{t-1} \right) - H^v \otimes x_2,$$

(16)

where \(H^c \in \mathbb{R}^{1 \times n \times n \times K}\) is a 4-D tensor stacked by \(\{h_k^c\}_{k=1}^K\). The solution of Eq. (13) can be represented as:

$$V^t = \text{prox}_{\eta_v \lambda_v}(V^{t-1} - \eta_v \nabla f(V^{t-1})),$$

(17)

where \(\text{prox}_{\eta_v \lambda_v}(\cdot)\) is the proximal operator correlated to the regularization term \(\psi_v(\cdot)\) with respect to \(V\).

**Updating C.** We first denote \(\tilde{x}_1 = x_1 - \sum_k d_k \otimes u_k\) and \(\tilde{x}_2 = x_2 - \sum_k h_k \otimes v_k\). Eq. (7) can be re-written as follows:

$$\min_C \frac{1}{2} \|\tilde{x}_1 - \sum_k d_k \otimes c_k\|_F^2 + \frac{1}{2} \|\tilde{x}_2 - \sum_k h_k \otimes c_k\|_F^2 + \lambda_C \psi_C(C),$$

(18)

Then the first two terms in Eq. (18) can be further combined and this yields the following optimization problem:

$$\min_C \frac{1}{2} \|M - \sum_k l_k \otimes c_k\|_F^2 + \lambda_C \psi_C(C),$$

(19)

where \(M\) is the concatenate of \(\tilde{x}_1\) and \(\tilde{x}_2\), \(l_k\) is the concatenation of \(d_k\) and \(h_k\). Thus, Eq. (19) can be solved by the same way as \(U\) and \(V\). Specifically, the quadratic approximation of this problem with respect to \(C\) is:

$$\min_C \frac{1}{2} \|C - (C^{t-1} - \eta_v \nabla f(C^{t-1}))\|_F^2 + \lambda_C \psi_C(C),$$

(20)

where

$$f(C^{t-1}) = \frac{1}{2} \|M - \sum_k l_k \otimes c_k\|_F^2,$$

(21)

$$\nabla f(C^{t-1}) = L^c \otimes \left( \sum_k (l_k \otimes c_k - M) \right),$$

(22)

where \(L^c \in \mathbb{R}^{1 \times n \times n \times K}\) is a 4-D tensor stacked by \(\{l_k^c\}_{k=1}^K\). Eq. (22) can be re-writed as:

$$\nabla f(C^{t-1}) = L^c \otimes \left( L^c \otimes C^{t-1} \right) - L^c \otimes M,$$

(23)

The solution of Eq. (20) can be represented as:

$$C^t = \text{prox}_{\eta_c \lambda_c}(C^{t-1} - \eta_v \nabla f(C^{t-1})), $$

(24)

where \(\text{prox}_{\eta_c \lambda_c}(\cdot)\) is the proximal operator correlated to the regularization term \(\psi_c(\cdot)\) with respect to \(C\).
3 Deep Unfolding MC-CDic Model

Before unfolding the above model, we first formulate some relaxations on the dictionaries. Following earlier works [Simon and Elad, 2019; Lecouat et al., 2020a; Lecouat et al., 2020b; Lecouat et al., 2020a; Liu et al., 2022], we untie the convolutional dictionaries from their transposed convolutional dictionaries in Eq. (11), (16) and (23) for better performance. To be specific, we initialize the convolution dictionary $D^0\otimes \bar{T}$ identically to $D^0\otimes T$ but are allowed to evolve independently during training. For convenience, we use $\tilde{D}^0\otimes T$ to replace $D^0\otimes T$. Similarly, $H^0\otimes T$ and $L^0\otimes T$ are also replaced by $H^0\otimes \bar{T}$ and $L^0\otimes \bar{T}$ in our deep unfolding model.

3.1 Network Design

(1) Feature Initialization Module. We use three convolutional dictionaries to initialize the common features and the unique features from the multi-contrast images $x_1$ and $x_2$. Specifically, we first use the dictionaries $\tilde{D}^0$ and $\tilde{H}^0$ to initialize the unique feature $U^0$ and $V^0$ respectively. Then we concatenate $\hat{x}_1$ and $x_2$, and use $L^0\bar{T}$ to initialize the common feature $C^0$.

(2) MC-CDic Iteration Module. As shown in the figure, our model consists of $T$ MC-CDic blocks, representing $T$ iterations of the algorithm for solving Eq. (4). Specifically, we take the input images $x_1$ and $x_2$, and the previous outputs $U^t$, $V^t$ and $C^t$ as inputs, and outputs the updated $U^{t+1}$, $V^{t+1}$ and $C^{t+1}$ in each iteration block. The iteration blocks are unfolded following the updating rules of Eq. (12), (17) and (24). To be specific, the convolutional dictionary operations can be easily implemented by the commonly-used convolutional layers in normal networks. The key issue of unfolding the algorithm is how to represent the three proximal operators $\operatorname{prox}_{\eta, \lambda_u}(\cdot)$, $\operatorname{prox}_{\eta, \lambda_c}(\cdot)$ and $\operatorname{prox}_{\eta, \lambda_v}(\cdot)$.

Inspired by recent unfolding networks [Wang et al., 2020; Wang et al., 2022], we employ the deep residual network (ResNet) to replace the proximal operators. Note that the $T$ MC-CDic iterations share the same parameters in our model. Thus, at the $t$th stage, the network for updating $U$ can be built as:

$$
\begin{align*}
\nabla f(U^{t+1}) &= \tilde{D}^0 \odot (D^0 \odot C^{t-1} + D^0 \odot U^{t-1}) - D^0 \odot x_1, \\
U^{t} &= \operatorname{prox}_{\eta_u}(U^{t-1} - \eta_u \nabla f(U^{t-1})),
\end{align*}
$$

where $\operatorname{prox}_{\eta_u}(\cdot)$ denotes the ResNet with several res-blocks. Similarly, the network for updating $V$ at the $t$th stage can be built as:

$$
\begin{align*}
\nabla f(V^{t+1}) &= \tilde{H}^0 \odot (H^0 \odot C^{t-1} + H^0 \odot V^{t-1}) - H^0 \odot \bar{x}_2, \\
V^{t} &= \operatorname{prox}_{\eta_v}(V^{t-1} - \eta_v \nabla f(V^{t-1})),
\end{align*}
$$

where $\operatorname{prox}_{\eta_v}(\cdot)$ denotes the ResNet with several res-blocks.

Based on the updated unique features $U^{t}$ and $V^{t}$, the network for updating $C$ at the $t$th stage can be built as:

$$
\begin{align*}
\nabla f(C^{t+1}) &= \tilde{L}^0 \odot (L^0 \odot C^{t-1} - \tilde{L}^0 \odot M^{t-1}) - \tilde{L}^0 \odot M^{t}, \\
C^{t} &= \operatorname{prox}_{\eta_c}(C^{t-1} - \eta_c \nabla f(C^{t-1})),
\end{align*}
$$

where $\operatorname{prox}_{\eta_c}(\cdot)$ denotes the ResNet with several res-blocks.

According to the updating rules of $U$, $V$ and $C$ in Eq. (25), (26) and (27), we design the multi-contrast convolutional dictionary (MC-CDic) block. Its overall structure is shown in Figure 2. All the parameters involved can be automatically learned from training data in an end-to-end manner.

3.3 Reconstruction Module. After $T$ MC-CDic iterations, we get the final common features $C^T$ and the unique features $U^T$ and $V^T$.

Figure 2: The overall structure of the proposed multi-contrast convolutional dictionary (MC-CDic) model for the guided-SR task.
feature $U^T$ and $V^T$. Then the reconstructed HR image $x'_2$ can be obtained by:

$$x'_2 = Q^c \otimes C^T + Q^v \otimes V^T,$$

where $P^c = \{p^c_k\}_{k=1}^K$ and $P^u = \{p^u_k\}_{k=1}^K$ are the common filters and unique filters for reconstructing the HR images.

### 3.2 Multi-Scale Dictionaries

The importance of multi-scale representations has been proven in numerous computer vision tasks. However, most existing CDic models only use the single-scale dictionary to extract single-scale representations, which limits their performance to a large extent. Inspired by Liu et al., 2022, we employ the multi-scale dictionaries in our MC-CDic model to perceive the multi-scale representations. To be specific, we employ a stripped-down version of the image synthesis process of U-Net by removing all non-linearities, batch normalization, and additive biases. We regard the decoder part of the U-Net as a normal CDic layer and regard the encoder part of the U-Net as its transposed CDic layer. As shown in Figure 3 (a) and (b), we present the operations of $U^t = \hat{D}^u \otimes s$ and $s = D^u \otimes U^t$. The proximal network for multi-scale dictionaries is shown in Figure 3 (c).

### 3.3 Loss Function

We employ the simple L1 loss to supervise the multi-contrast images simultaneously. It can be formulated as:

$$\mathcal{L} = \alpha ||x'_1 - \hat{x}_1||_1 + ||x'_2 - \hat{x}_2||_1,$$

where $\alpha$ is a trade-off parameter to balance the importance of different contrasts. In this paper, we set $\alpha = 0.1$. Note that the loss on the reference image is to ensure that the reference image can be efficiently decomposed.

### 4 Experiments

#### 4.1 Datasets

Following Lyu et al., 2020; Fang et al., 2022, we employ two publicly available multi-modal MR image datasets, IXI and BraTS2018 [Menze et al., 2015] to validate the effectiveness of our model. IXI dataset contains 576 multi-contrast MR volumes and we employ the PD as the reference modal to guide the reconstruction of the target T2 modal. BraTS2018 dataset contains 285 multi-contrast MR volumes and we employ the T1 as the reference modal to guide the reconstruction of the target T2 modal. We splitted the two datasets patient-wise into a ratio of 7:1:2 for training/validation/testing. For each volume, we only select the middle 100 slices for our experiments. We use cartesian sampling mask and the k-space center cropping [Lyu et al., 2020] mask to get the undersampled and LR images, respectively. The acceleration ratio for reconstruction and the scale factor for SR are set at 4.

#### 4.2 Implementation Details

**Training Details.** Our proposed MC-CDic model is implemented in PyTorch with an NVIDIA RTX3090 GPU. We adopt the Adam optimizer with a batch size of 6. The learning rate is set to $1 \times 10^{-4}$ and the models are trained 50 epochs. We employ the peak-to-noise-ratio (PSNR), structural similarity index (SSIM), and root mean squared error (RMSE) to evaluate the model performance. The higher PSNR and SSIM, the lower RMSE represent the better result.

**Model Details.** In our large model MC-CDic-L, the iteration stage $T$ is set to 4. The levels of the multi-scale dictionaries are set to 3. From the first level to the third level, the channel numbers of the features are set to 64, 96, and 128. The numbers of the resblocks for the multi-scale proximal network are set to 5, 3, and 1. In our small model MC-CDic-S, we use the single-scale dictionary and other settings are the same with MC-CDic-L.

#### 4.3 Comparison with State-of-the-arts

**Evaluations on Guided-SR.** Guided-SR aims to reconstruct the HR image from its LR image with the aid of another reference HR image. We compare our methods with
Figure 4: Error maps of different single-contrast SR (the first three methods) and guided-SR (the next six methods) with the scale factor of 4 on the IXI (PD guides T2) and BrainTS (T1 guides T2) testing sets.

Figure 5: Visualization of the decomposed common and unique components of our MC-CDic model on the guided-SR task.

Table 2: Quantitative results of single-contrast MRI reconstruction and multi-contrast guided-reconstruction algorithms at 4 acceleration.

4.4 Ablation Analysis

The Effect of the Number of Stages $T$. To show how the number of iteration stages $T$ affects the reconstruction performance, we have compared the proposed MC-CDic method with different stages. Figure 7 shows the average PSNR and SSIM performance for the guided-SR task and guided-reconstruction task on the IXI validation set when $T \in [1, 5]$. It can be observed that more stages lead to improved reconstruction performance. Specifically, from $T = 1$ to $T = 4$, the performance improves significantly.
Figure 6: Visual comparison of different single-contrast reconstruction (the first four methods) and multi-contrast guided reconstruction (the last five methods) on the IXI (PD guides T2) and BrainTS (T1 guides T2) testing sets when acceleration is $\times 4$.

Figure 7: The PSNR and SSIM curves on IXI validation set with a different number of stages $T$.

Figure 8: PSNR and SSIM comparisons on IXI dataset with different levels of multi-scale dictionaries.

<table>
<thead>
<tr>
<th>Method</th>
<th>Guided-SR</th>
<th>Guided-reconstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>SSIM</td>
<td>RMSE$^1$</td>
</tr>
<tr>
<td>w/ proxNet</td>
<td>42.15</td>
<td>0.9799</td>
</tr>
<tr>
<td>w proxNet</td>
<td>42.58</td>
<td>0.9810</td>
</tr>
</tbody>
</table>

Table 3: Quantitative results of the MC-CDic model with and without proximal network on IXI testing set.

**The Effect of the Multi-Scale Dictionaries.** We compare the performance of the proposed MC-CDic model under different levels of multi-scale dictionaries in Figure 8. From the figure, we can find that the performance increases with the feature levels. From the trend lines, we can also see that multi-scale features are more important for the reconstruction task. Specifically, three levels of feature improve the PSNR by about 0.5dB and 0.9dB compared with the single-level features on the SR and reconstruction tasks, respectively. More feature levels may achieve better performance, however, the number of parameters and computations will also increase. To balance the performance and the computational complexity, we manually set $T = 4$ in our final model.

**The Effect of the Proximal Network.** In this paper, we employ a proximal network to learn the prior information in a data-driven manner. To validate its effectiveness, we compare the performance of the proposed MC-CDic model with and without the proximal network. Note that when we remove the proximal network, we use the soft-thresholding operator to constrain the learned features. The performance of these two models is shown in Table 3. For the guided-SR task, the proximal network improves the performance by 0.43dB PSNR on the IXI-T2 dataset. For the guided-reconstruction task, the proximal network improves the performance by 1.06dB PSNR on the IXI-T2 dataset. It fully demonstrates the effectiveness of the proposed proximal network.

**5 Concluding Remarks**

We proposed a multi-scale convolutional dictionary model (MC-CDic) for multi-contrast MRI reconstruction and SR. Different from existing DCNN-based methods that manually designed fusion rules to fuse the multi-contrast information, our MC-CDic model was constructed under the guidance of the optimization algorithm with a well-designed data fidelity term. A proximal gradient algorithm was employed to solve the model and the iterative solutions were unfolded into a deep network. Especially, each module in the network could one-to-one correspond to an operation of a solution, thus our model was interpretable. We have tested our MC-CDic model on multi-contrast MRI reconstruction and SR tasks, experimental results demonstrate the superior performance of our method over the existing SOTA methods. In the future, we will explore more applications of our model on multi-modal tasks, such as RGB-guided depth map completion and RGB-guided depth map SR.
Acknowledgements
This work was supported by the National Key R&D Program of China (2022ZD0161800), the NSFC-RGC (61961160734), the Shanghai Rising-Star Program (21QA1402500), and the Shanghai Municipal Commission of Economy and Informatization Program (2021-GZL-RGZN-01028).

References


