

# Decomposition-Based Variational Network for Multi-Contrast MRI Super-Resolution and Reconstruction

Pengcheng Lei<sup>1</sup> Faming Fang<sup>1,2\*</sup> Guixu Zhang<sup>1</sup> Tiejong Zeng<sup>3</sup>

<sup>1</sup>School of Computer Science and Technology, East China Normal University

<sup>2</sup>KLATASDS-MOE <sup>3</sup>Department of Mathematics, Chinese University of Hong Kong

pengchenglei1995@163.com {fmfang, gxzhang}@cs.ecnu.edu.cn zeng@math.cuhk.edu.hk

## Abstract

Multi-contrast MRI super-resolution (SR) and reconstruction methods aim to explore complementary information from the reference image to help the reconstruction of the target image. Existing deep learning-based methods usually manually design fusion rules to aggregate the multi-contrast images, fail to model their correlations accurately and lack certain interpretations. Against these issues, we propose a multi-contrast variational network (MC-VarNet) to explicitly model the relationship of multi-contrast images. Our model is constructed based on an intuitive motivation that multi-contrast images have consistent (edges and structures) and inconsistent (contrast) information. We thus build a model to reconstruct the target image and decompose the reference image as a common component and a unique component. In the feature interaction phase, only the common component is transferred to the target image. We solve the variational model and unfold the iterative solutions into a deep network. Hence, the proposed method combines the good interpretability of model-based methods with the powerful representation ability of deep learning-based methods. Experimental results on the multi-contrast MRI reconstruction and SR demonstrate the effectiveness of the proposed model. Especially, since we explicitly model the multi-contrast images, our model is more robust to the reference images with noises and large inconsistent structures. The code is available at <https://github.com/lpccc-cv/MC-VarNet>.

## 1. Introduction

Magnetic Resonance Imaging (MRI) is a noninvasive and non-ionizing medical imaging technique, which has been widely used in medical diagnosis, clinical analysis, and staging of disease [27]. However, MRI scanning is always time-consuming due to physics and physiological

\*Corresponding author

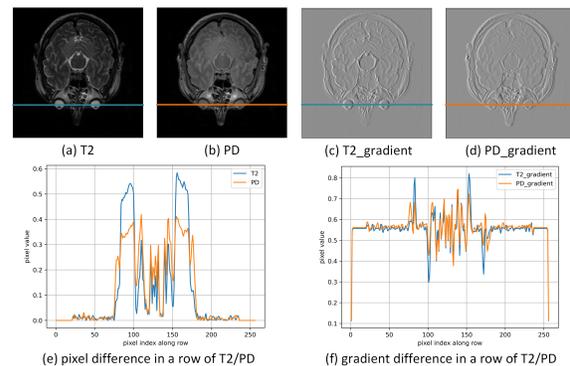


Figure 1. The visualization of the multi-contrast images, i.e. T2 and PD, on IXI dataset. They have their unique contrast information (as shown in (e)), and they also share some common structure information (as shown in (f)).

constraints. The scanning time of each patient usually takes more than ten minutes, which will affect the efficiency of diagnosis and cause discomfort to patients [8]. To accelerate the acquisition of MRI, it is common to acquire the undersampled  $k$ -space measurements and then use the post-processing algorithms (e.g. super-resolution (SR) and reconstruction) to restore its fully-sampled one.

The purpose of the MRI reconstruction is to eliminate aliasing artifacts caused by  $k$ -space undersampling and the MRI SR aims to restore the missing high-frequency details from the low-resolution (LR) ones. In recent years, numerous MRI SR [44, 31, 42] and reconstruction [34, 29, 45, 23] methods have been emerged. However, they only use single-contrast MR images and fail to utilize the information from other modalities. In fact, radiologists tend to acquire images of bodies with different contrasts (e.g. T1-weighted and T2-weighted) to comprehensively evaluate patients' conditions. Besides, since the sampling time of different contrasts is quite different, it is promising to utilize an HR (or fully-sampled) reference image with a shorter acquisition time to reconstruct the image with a longer scanning time [14]. We call them multi-contrast reconstruction

and multi-contrast SR.

In recent years, deep convolution neural network (DCNN) has become a mainstream approach for solving the multi-contrast reconstruction [36, 30, 1, 40, 8] and multi-contrast SR [20, 7, 14, 6] problems. By designing proper fusion rules, reference information can be effectively transferred to the target image. Despite the promising results of them, their networks are always handcrafted black boxes, which lack certain interpretations. Model-based methods [11, 35, 28] construct optimization models by designing priors for specific problems, and then they are solved by optimization algorithms, which are more interpretable than black-box DCNN methods. To incorporate interpretability and domain knowledge into deep networks, the model-driven DCNN methods have emerged to tackle various problems of image processing, including deraining [33], deblurring [17], super-resolution [2, 3], CT metal artifact reduction [32] and so on. Their success also inspired us to design an interpretable and powerful model for the multi-contrast SR and reconstruction tasks.

In this paper, we propose a deep unfolding multi-contrast variational network (MC-VarNet) to explicitly model the correlations of multi-contrast images. According to our observation, as shown in Figure 1, multi-contrast images have the following two characteristics. Firstly, since they are sampled from the same body part, they usually contain some common structural information. Secondly, since the two images are sampled under different scan settings, they have their unique contrast information and some inconsistent structures. To effectively transfer useful information to the target image as well as avoid the interference of inconsistent information, we build a variational model to reconstruct the target image and decompose the reference image into a common component and a unique component. In the feature interaction phase, only the common component is transferred to the target image.

To solve the variational model, we employ the half-quadratic splitting algorithm to optimize each variable and construct an MC-VarNet by unfolding the iterative steps into deep neural networks. Similar to learning-based methods, MC-VarNet adaptively learns the deep priors from the reconstructed images and the decomposed components in a data-driven manner. Similar to model-based methods, MC-VarNet explicitly models the correlations of multi-contrast images, which is more transparent. The proposed MC-VarNet thus combines the advantages of model-based and deep learning-based methods. Our contributions can be summarized as:

- We propose a variational model to solve the multi-contrast MRI SR and reconstruction problems simultaneously. In our model, we explicitly model the relationship of multi-contrast images based on an observation that different contrast images have consistent and

inconsistent information.

- We optimize the variational model using the half-quadratic splitting algorithm and unfold the iterative step into deep networks. Thus our model combines the good interpretability of the model-based methods with the powerful feature expression ability of the deep neural networks.
- We test our model on guided MRI SR and reconstruction tasks. Experiments demonstrate the effectiveness of the proposed model. Besides, since we explicitly model the multi-contrast images, the proposed model is more robust to the noise-polluted and inconsistent reference images compared with existing methods.

## 2. Related work

### 2.1. Single-contrast MRI SR and reconstruction

DCNN has been the mainstream solution for solving the MRI SR and reconstruction problems. Qui et al. [24] designed a DCNN model for knee MRI SR. Lyu et al. [21] employed ensemble learning to improve the performance of a single model. Zhao et al. [44] design a channel-splitting network to fully utilize the hierarchical features. Li et al. [15] used attention mechanisms for pelvic image SR. Zhang et al. [42] proposed a squeeze and excitation reasoning attention network to sense the entire space of the MR images. For MRI reconstruction, Wang et al. [34] first proposed a multi-layer CNN model for MRI reconstruction. Jin et al. [12] and Yang et al. [38] proposed model-based unrolling methods by combining prior regularization. [29, 9, 4] proposed variational networks to solve the single-contrast MRI reconstruction problem. To further improve the model performance, the cross-domain methods [5, 23, 46] were proposed. These methods have got satisfactory results, however, they all focus on reconstructing images by using single-contrast MR images and fail to utilize the multi-contrast information.

### 2.2. Multi-contrast MRI SR and reconstruction

Multi-contrast imaging is a main feature of MR images. For solving the multi-contrast MRI SR task, Lyu et al. [20] fused the multi-contrast information in the high-level feature space. Feng et al. [7] designed a multi-stage integration network to fuse the multi-contrast features at different stages. Li et al. [14] proposed a transformer-empowered multi-scale contextual matching network to capture long-range dependencies for MR images. Fang et al. [6] proposed the cross-modality transformer network to explore valuable prior knowledge from multi-contrast images. For the multi-contrast MRI reconstruction task, Xiang and Dar et al. [36, 1] simply concatenated the multi-contrast images as their model inputs. Sun et al. [30] sent the multi-contrast

images into the network together and restore them simultaneously. Feng et al. [8] proposed a multi-modal transformer to transform the features from the reference contrast to the target contrast. Although they have achieved good reconstruction results, they fuse the multi-contrast information by manually designing fusion rules, which lack sufficient interpretability, thus limiting further performance improvements.

### 3. Variational model for multi-contrast MRI

#### 3.1. Degradation model

The problem is to reconstruct HR (or fully-sampled) images from under-sampled  $k$ -space data. The acquisition process of the under-sampled data can be written as:

$$k_x = M \odot \mathcal{F}(\hat{x}) + \varepsilon. \quad (1)$$

Here  $\hat{x} \in \mathbb{R}^{m \times n}$  is the fully-sampled MR image,  $k_x \in \mathbb{R}^{m \times n}$  is the under-sampled  $k$ -space data what we actually observed,  $\varepsilon$  is the measurement noise, and  $\mathcal{F}(\cdot)$  is Fourier Transform operator. The values of  $M \in \mathbb{R}^{m \times n}$  stand for the corresponding  $k$ -space positions are sampled or not.  $\odot$  is the pixel-wise multiply operation. For the reconstruction task,  $M$  represents the cartesian sampling mask [8]. For the MRI SR task,  $M$  is the center cropping mask [44, 14], where only the central low-frequency data are kept and the peripheral areas are set to zero. The under-sampled MR image  $\tilde{x}$  can be obtained by  $\mathcal{F}^{-1}(k_x)$ , where  $\mathcal{F}^{-1}(\cdot)$  represents inverse Fourier Transform.

#### 3.2. Model formulation

Let  $\hat{x}_1 \in \mathbb{R}^{m \times n}$  and  $\hat{x}_2 \in \mathbb{R}^{m \times n}$  represent two HR (or fully-sampled) MR images with different contrasts (e.g. T2 and PD). The LR (or under-sampled) image of  $\hat{x}_1$  is  $\tilde{x}_1$ . Here we regard  $\hat{x}_1$  as the target modality and regard  $\hat{x}_2$  as the reference modality. Take guided-SR for example, our model aims to restore the HR image  $x_1$  from LR image  $\tilde{x}_1$  with the aid of  $\hat{x}_2$ . As aforementioned, multi-contrast images are sampled under the same anatomical structure, they usually share some common structure information. Since they are sampled with different scan settings, they have their unique contrast information. Based on such observations, we build the variational model:

$$\begin{aligned} \operatorname{argmin}_{\{x_1, C, U\}} & \frac{1}{2} \|M\mathcal{F}(x_1) - k_{x_1}\|_2^2 + \frac{1}{2} \|C + U - \hat{x}_2\|_2^2 \\ & + \frac{\gamma}{2} \|Ax_1 - BC\|_2^2 + \lambda_1 \mathcal{R}(x_1) \\ & + \lambda_2 \psi_2(C) + \lambda_3 \psi_3(U), \end{aligned} \quad (2)$$

where the first term in Eq. (2) is the data fidelity for the target modality that ensures consistency between the reconstructed image  $x_1$  and the under-sampled  $k$ -space data

$k_{x_1}$ , the second term is the data fidelity for the reference modality that decompose it into common component  $C$  and unique component  $U$ .  $A$  and  $B$  are two transformation matrices to transform the two images to the feature domain, as well as constrain their similarity in the feature domain.  $\gamma$ ,  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  are the trade-off parameters.  $\mathcal{R}(\cdot)$ ,  $\psi_2(\cdot)$ ,  $\psi_3(\cdot)$  are the regularization terms.

#### 3.3. Optimization algorithm

Half-quadratic splitting (HQS) algorithm is employed to solve Eq. (2). To be specific, we introduce auxiliary variables  $P, Q$  to replace  $C$  and  $U$ . Eq. (2) can be reformulated as:

$$\begin{aligned} \operatorname{argmin}_{\{x_1, C, U, P, Q\}} & \frac{1}{2} \|M\mathcal{F}(x_1) - k_{x_1}\|_2^2 + \frac{1}{2} \|C + U - \hat{x}_2\|_2^2 \\ & + \frac{\gamma}{2} \|Ax_1 - BC\|_2^2 + \frac{\alpha}{2} \|P - C\|_2^2 + \frac{\beta}{2} \|Q - U\|_2^2 \\ & + \lambda_1 \mathcal{R}(x_1) + \lambda_2 \psi_2(P) + \lambda_3 \psi_3(Q), \end{aligned} \quad (3)$$

where  $\alpha$  and  $\beta$  are hyper-parameters. Eq. (3) can be solved via the following iterative scheme:

$$x_1^{t+1} = \operatorname{argmin}_{x_1} \frac{1}{2} \|M\mathcal{F}(x_1) - k_{x_1}\|_2^2 + \frac{\gamma^t}{2} \|Ax_1 - BC^t\|_2^2 + \lambda_1 \mathcal{R}(x_1), \quad (4)$$

$$\begin{cases} C^{t+1} = \operatorname{argmin}_C \frac{1}{2} \|C + U^t - \hat{x}_2\|_2^2 + \frac{\gamma^t}{2} \|Ax_1^t - BC\|_2^2 \\ \quad + \frac{\alpha^t}{2} \|P^t - C\|_2^2, \\ U^{t+1} = \operatorname{argmin}_U \frac{1}{2} \|C^t + U - \hat{x}_2\|_2^2 + \frac{\beta^t}{2} \|Q^t - U\|_2^2, \end{cases} \quad (5)$$

$$\begin{cases} P^{t+1} = \operatorname{argmin}_P \frac{\alpha}{2} \|P - C^{t+1}\|_2^2 + \lambda_2 \psi_2(P), \\ Q^{t+1} = \operatorname{argmin}_Q \frac{\beta}{2} \|Q - U^{t+1}\|_2^2 + \lambda_3 \psi_3(Q), \end{cases} \quad (6)$$

where  $t$  is the number of iterations. Next, we will solve these sub-problems separately.

**Solving  $x_1$ :** Eq. (4) can be solved by gradient descent algorithm [29] and the image  $x_1^{t+1}$  can be updated from  $x_1^t$  using:

$$\begin{aligned} x_1^{t+1} = & x_1^t - \mu^t F^*(F(x_1^t) - k_{x_1}) \\ & - \mu^t (\gamma^t A^\top (Ax_1^t - BC^t) + \lambda_1 \phi(x_1^t)). \end{aligned} \quad (7)$$

Here  $\mu^t$  is the learning rate,  $F(\cdot)$  is the linear forward operator that applies 2D Fourier Transform and then under-samples the data using sampling mask  $M$ ,  $\phi(x_1^t)$  is the gradient operator of  $\mathcal{R}$  with respect to  $x_1$ ,  $F^*(\cdot)$  is the hermitian of the forward operator  $F(\cdot)$  [29].

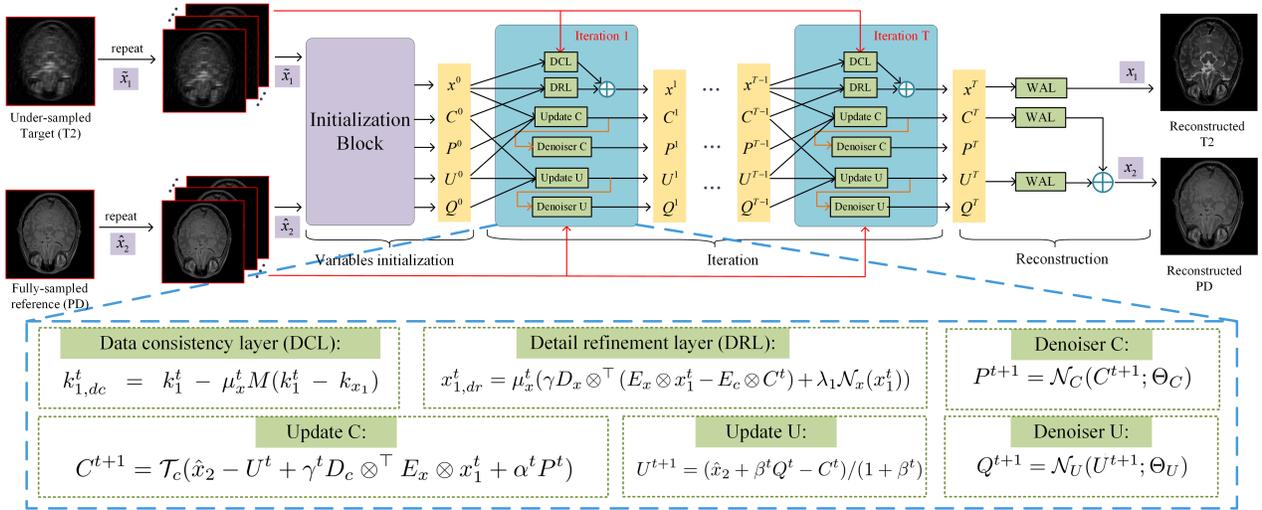


Figure 2. The overall structure of the proposed multi-contrast variational network for the multi-contrast guided-SR task.

**Solving  $C$  and  $U$ :** The sub-problem (5) can be directly solved. Thus we can get the closed-form solution of  $C$  and  $U$  as follows:

$$\begin{cases} C^{t+1} = (I + \gamma^t B^\top B + \alpha^t I)^{-1} (\hat{x}_2 - U^t \\ \quad + \gamma^t B^\top A x_1^t + \alpha^t P^t), \\ U^{t+1} = (\hat{x}_2 + \beta^t Q^t - C^t) / (1 + \beta^t), \end{cases} \quad (8)$$

where  $I$  is the identity matrix. In the above iteration steps, we estimate the common component  $C$  and unique component  $U$  simultaneously. The key problem is how to define the implicit priors and how to solve the sub-problem (6). Inspired by recent model-based deep unfolding methods [41, 4, 37], we regard sub-problems (6) as denoising problems for  $C$  and  $U$ . Thus deep priors can be extracted from these networks and Eq. (6) can be described as follows:

$$\begin{cases} P^{t+1} = \mathcal{N}_C(C^{t+1}; \Theta_C), \\ Q^{t+1} = \mathcal{N}_U(U^{t+1}; \Theta_U), \end{cases} \quad (9)$$

where  $\mathcal{N}_C, \mathcal{N}_U$  are the deep networks for the common component and unique component, respectively.  $\Theta_C, \Theta_U$  are the parameters of the denoising networks.

## 4. Deep unfolding multi-contrast network

Based on the above analysis, we will construct our deep unfolding multi-contrast variational network following the updating rules of Eq.(7), (8) and (9). As shown in Figure 2, the proposed model contains three submodules: variables initialization module, iteration module and reconstruction module.

### 4.1. Initialization module

Since the original MR image only has one grey channel, in this paper, we expand the channel number to allow

more diverse information for reconstructing the target image [25, 32]. Specifically, we copy each image along the channel dimension to expand the channel number from 1 to  $c_i$ . In the iteration stages, our model reconstructs  $c_i$  images simultaneously. To initialize the five variables that appeared in our model, we first employ two denoising networks  $\mathcal{N}_x$  and  $\mathcal{N}_c$  to initialize  $x_1^0$  and  $C^0$ . Then  $U^0$  is initialized as  $\hat{x}_2 - C^0$ ,  $P^0$  and  $Q^0$  are initialized as  $C_0$  and  $U_0$  respectively.

### 4.2. Iteration module

The structure of the proposed model is shown in Figure 2. It contains of  $T$  iteration blocks, representing  $T$  iterations of the algorithm for solving Eq. (2). To be specific, we regard the input images  $\hat{x}_1$  and  $\hat{x}_2$ , and the previous outputs  $x_1^t, C^t, U^t, P^t$  and  $Q^t$  as inputs, and outputs the updated  $x_1^{t+1}, C^{t+1}, U^{t+1}, P^{t+1}$  and  $Q^{t+1}$  in each iteration block. The iteration blocks are designed following the updating rules of Eq. (7), (8) and (9). Next, we will introduce the design details of each deep unfolding module.

**Modules for updating  $x_1$ :** Eq. (7) has shown how to update variable  $x_1$  from  $x_1^t$  to  $x_1^{t+1}$ . In practice, there are two key problems to be solved. The first problem is how to define implicit prior term  $\phi(x_1^t)$ . Inspired by existing variational networks [9, 29], we directly use a DCNN to learn the prior information. The second problem is the implementation of the modality transformation term  $A^\top (A x_1^t - B C^t)$ , which transforms the information from the reference common component  $C^t$  to target image  $x_1^{t+1}$ . Inspired by recent convolutional dictionary (CDic) learning methods [3, 18], we implement the transformation operators as CDic layers.

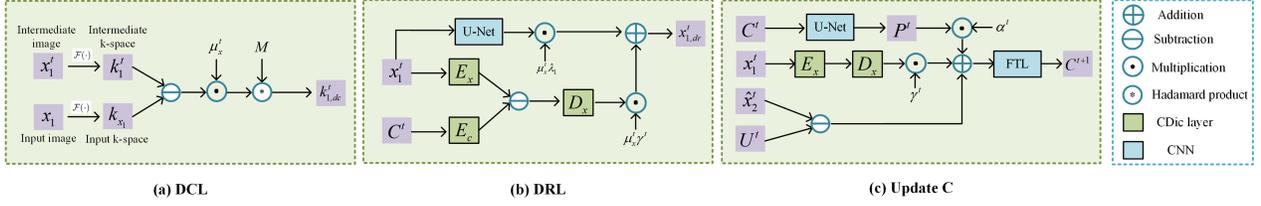


Figure 3. The structure of the proposed submodules of the proposed MC-VarNet. (a) Data consistency layer (DCL), (b) detail refinement layer (DRL), and (c) submodule for updating  $C$ .

Thus, Eq. (7) can be reformulated as:

$$x_1^{t+1} = x_1^t - \mu_x^t F^*(F(x_1^t) - k_{x_1}) + \mu_x^t (\gamma^t D_x \otimes^\top (E_x \otimes x_1^t - E_c \otimes C^t) + \lambda_1 \mathcal{N}_x(x_1^t)), \quad (10)$$

where  $\otimes$  denotes the convolutional operation,  $\otimes^\top$  denotes the transposed convolution.  $E_x \in \mathbb{R}^{c_f \times 3 \times 3 \times c_i}$  and  $E_c \in \mathbb{R}^{c_f \times 3 \times 3 \times c_i}$  are the encoder convolutional dictionary layers to transform images  $x_1^t$  and  $C^t$  to the feature domain.  $D_x \in \mathbb{R}^{c_i \times 3 \times 3 \times c_f}$  is the decoder convolutional dictionary layer to transform features to the image domain. Note that  $c_i$  and  $c_f$  are the channel numbers of the input image and the intermediate feature, respectively.  $\mathcal{N}_x(\cdot)$  is a modified UNet [26], which is used to learn the implicit priors from  $x_1^t$ .

Similar to [29], Eq. (10) can be regarded as a data-consistency layer (DCL) in the  $k$ -space and a detail refinement layer (DRL) in the image domain, which can be reformulated as:

$$x_1^{t+1} = \mathcal{F}^{-1}(k_{1,dc}^t) + x_{1,dr}^t, \quad (11)$$

where  $k_{1,dc}^t = k_1^t - \mu_x^t M(k_1^t - k_{x_1})$ , and  $x_{1,dr}^t = \mu_x^t (\gamma^t D_x \otimes^\top (E_x \otimes x_1^t - E_c \otimes C^t) + \lambda_1 \mathcal{N}_x(x_1^t))$ . Here  $k_{1,dc}^t$  is the reconstructed  $k$ -space data after DCL,  $x_{1,dr}^t$  is the refinement details learned by DRL,  $k_1^t$  is the intermediate  $k$ -space data and  $k_{x_1} = \mathcal{F}(x_1)$ ,  $k_{x_1}$  is the original input  $k$ -space data and  $k_{x_1} = \mathcal{F}(\tilde{x}_1)$ .

In Figure 3, we show the network structures of the DCL and DRL. DCL enforces the  $k$ -space data-consistency between the intermediate reconstruction images and the original input images. DRL has two branches, one branch employs a U-Net to learn the deep priors from the target image, and the other branch transfers the complementary information from the reference image to the target image. Adding the results of the two branches, we can get the refinement details of the target image.

**Modules for updating  $C$  and  $P$ :** For updating  $C$ , we first introduce the CDic layers to implement the transformation functions in Eq. (8) and then simplify the operation  $(I + \gamma^t B^\top B + \alpha^t I)^{-1}$  as a simple feature transformation layer (FTL). The equation for updating  $C$  and  $P$  can be for-

mulated as:

$$\begin{cases} C^{t+1} = \mathcal{T}_c(\hat{x}_2 - U^t + \gamma^t D_c \otimes^\top E_x \otimes x_1^t + \alpha^t P^t), \\ P^{t+1} = \mathcal{N}_C(C^{t+1}; \Theta_C), \end{cases} \quad (12)$$

$\mathcal{N}_C(\cdot)$  is implemented as a deep denoising network for updating auxiliary variable  $P$ .  $\mathcal{T}(\cdot)$  denotes the FTL, which contains two ‘‘Conv’’ layers and a ‘‘ReLU’’ activation layer. The network structure for updating  $C$  is shown in Figure 3. It contains three branches. The first branch employs a denoising network to suppress the noises of  $C^t$ . The second branch employs an encoder CDic layer and a decoder CDic layer to transfer the information from the target image to the common component of the reference image. The third branch is the decomposition-based data fidelity term. Finally, a simple FTL is employed to get the updated  $C$ .

**Modules for updating  $U$  and  $Q$ :** Variables  $U$  and  $Q$  can be directly updated by:

$$\begin{cases} U^{t+1} = (\hat{x}_2 + \beta^t Q^t - C^t) / (1 + \beta^t), \\ Q^{t+1} = \mathcal{N}_U(U^{t+1}; \Theta_U). \end{cases} \quad (13)$$

**Denoising network:** We design a modified U-Net as our deep denoising network to update the estimated  $P$  and  $Q$  in Eq. (12) and (13). Our denoising U-Net consists of three encoder blocks and three decoder blocks. Each block contains several ‘‘Conv’’ layers and ‘‘ReLU’’ layers. Besides, skip connections are used to fuse the information between the encoders and decoders. It should be noticed that the denoising networks in different iteration stages share the same network parameters in our model. More details about the denoising network can be found in our supplementary material.

### 4.3. Reconstruction module

After  $T$  iterations, we have got the reconstruction results of the target image  $x_1^T$  and the decomposed common components  $C^T$  and unique components  $U^T$ . As depicted above, we reconstruct  $c_i$  images in the iteration stages simultaneously. Here we employ a weighted average layer (WAL) to get the final reconstruction result. The final result

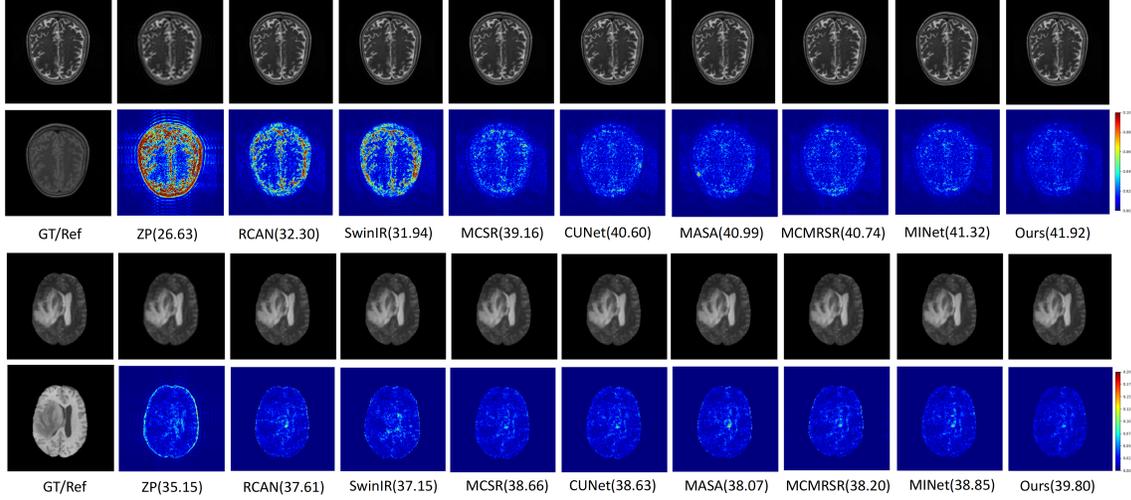


Figure 4. Visual comparison of different single-contrast SR (ZP, RCAN [43] and SwinIR [16]) and multi-contrast guided SR (MCSR [20], CUNet [3], MASA [19], MCMRSR [14] and MINet [7]) on the IXI and BrainTS testing sets when the scaling factor is  $\times 4$ .

can be obtained by:

$$\begin{cases} x_1 = \frac{1}{c_i} \sum_{i=1}^{c_i} w_1^i x_1^T(i), \\ x_2 = \frac{1}{c_i} \sum_{i=1}^{c_i} w_2^i (C^T(i) + U^T(i)), \end{cases} \quad (14)$$

where  $w_1^i$  and  $w_2^i$  are the learnable weights for reconstructing  $x_1$  and  $x_2$ , respectively.

#### 4.4. Loss function

The L1 loss is used to supervise the reconstruction results, which can be represented as:

$$\mathcal{L} = \eta_1 \|x_1 - \hat{x}_1\|_1 + \eta_2 \|x_2 - \hat{x}_2\|_1, \quad (15)$$

where  $\eta_1$  and  $\eta_2$  are the hyper-parameters to weigh the importance of different contrast images. Here we set  $\eta = 1$  and  $\eta_2 = 0.1$ . The loss on the reference image is for better decomposition.

## 5. Experiments

### 5.1. Datasets and implementation details

**Datasets.** Following [20, 6], two publicly available multi-modal MR image datasets (IXI and BraTS2018 [22]) is used to evaluate the performance of the proposed model. The IXI dataset contains 576 multi-contrast MR volumes. In our experiments, the PD modality is used as the reference image to guide the target T2 modality. The BraTS2018 dataset has 285 multi-contrast MR volumes. We regard the T1 modality as the reference image to guide the reconstruction of the target T2 modality. The two datasets are splitted

into the training, validation, and testing sets with a ratio of 7:1:2. Note that only the middle 100 slices are used in our experiments. In the following experiments, we set the acceleration ratio (multi-contrast reconstruction task) and the scale factor (multi-contrast SR task) to 4.

**Training details.** Our model is constructed based on PyTorch with two NVIDIA RTX3090 GPUs. The Adam optimizer is used to optimize the model. In the training phase, the batch size is set to 6 and the learning rate is set to  $1 \times 10^{-4}$ . The models are trained 50 epochs on the training data. The peak-to-noise-ratio (PSNR), structural similarity index (SSIM), and root mean squared error (RMSE) are employed to evaluate the model performance. The higher PSNR and SSIM, the lower RMSE indicates the better result.

**Model details.** In the large version MC-VarNet-L, the iteration stages  $T$  are set to 4. The input channel  $c_i$  and the intermediate feature channel  $c_f$  are set to 64. In the small version MC-VarNet-S, the iteration stages  $T$  are set to 4.  $c_i$  and  $c_f$  are set to 32.

### 5.2. Comparison with state-of-the-arts

**Evaluations on multi-contrast MRI SR.** We compare our models with various single image SR approaches, including zero padding (ZP), VDSR [13], CSN [44], RCAN [43] and SwinIR [16], and some existing SOTA multi-contrast MRI SR methods, including MCSR [20], CUNet [3], MASA [19], MINet [7] and MCMRSR [14]. The qualitative results of these methods are shown in Table 1. From the table, we can find that the proposed MC-VarNet-L model achieves the best performance. Specifically, it outperforms the second-place MINet by 0.55dB and 0.8dB in PSNR on the IXI and BrainTS datasets, respectively. The proposed MC-VarNet-S has only 1.4M pa-

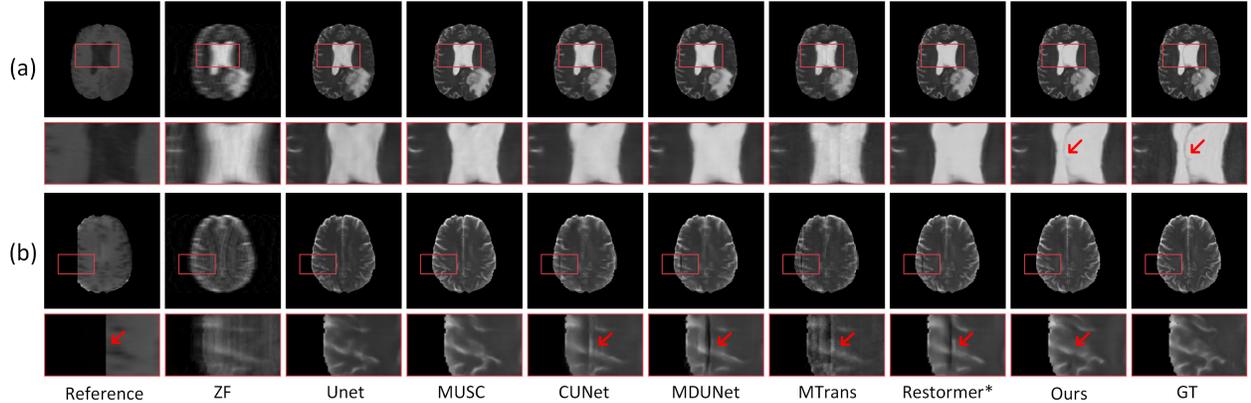


Figure 5. Visual comparison of different single-contrast reconstruction (ZF, UNet [26] and MUSC [18]) and multi-contrast guided reconstruction (CUNet [3], MDUNet [36], MTrans [8] and Restormer\* [39]) on the BrainTS testing sets when acceleration is  $\times 4$ .

Table 1. Quantitative results of different single-contrast MRI SR and multi-contrast MRI guided-SR algorithms when scaling factor is  $\times 4$ .

Method	Params	IXI-T2			BrainTS-T2		
		PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$
ZP	-	30.43	0.8840	8.15	32.40	0.8873	6.38
VDSR [13]	0.7M	32.51	0.9162	6.37	35.53	0.9580	4.46
CSN [44]	11.2M	34.04	0.9365	5.35	36.30	0.9633	4.08
RCAN [43]	15.6M	34.44	0.9413	5.12	36.81	0.9670	3.84
SwinIR [16]	11.9M	34.37	0.9402	5.17	36.99	0.9680	3.77
CUNet [3]	0.2M	39.65	0.9720	2.77	38.17	0.9710	3.32
MCSR [20]	3.5M	40.00	0.9734	2.67	38.09	0.9718	3.36
MASA [19]	4.0M	41.45	0.9786	2.23	37.66	0.9700	3.50
MCMRSR [14]	3.5M	40.93	0.9753	2.53	37.82	0.9660	3.45
MINet [7]	11.9M	41.70	0.9784	2.22	39.46	0.9758	2.88
Ours-S	1.4M	<b>41.71</b>	<b>0.9787</b>	<b>2.21</b>	<b>39.81</b>	<b>0.9775</b>	<b>2.78</b>
Ours-L	5.7M	<b>42.25</b>	<b>0.9787</b>	<b>2.08</b>	<b>40.26</b>	<b>0.9791</b>	<b>2.65</b>

rameters, but it performs better than most existing methods. In Figure 4, We visualize the reconstruction error maps of different methods on two images. It can be clearly seen that our methods have fewer reconstruction errors than other methods.

### Evaluations on multi-contrast MRI reconstruction.

We compare our models with various single-contrast reconstruction methods, including zero filling (ZF), UNet [26], MUSC [18] and Restormer [39], and some SOTA multi-contrast guided-reconstruction methods, including MDUNet [36], MTrans [8] and Restormer\* [39]. Note that we concatenate the multi-contrast images as the input of the Restormer\* to restore the target image. We show the results in Table 2. As listed in the table, our MC-VarNet-L and MC-VarNet-S achieve the best and the second-best performance on the two testing sets. Figure 5 (a) shows the visual comparisons of these methods. Among them, our method restores more anatomical details than other methods.

**Robustness analysis.** We manually add Gaussian noise  $\sigma = (5, 7, 9, 11, 13, 15)$  in the reference images when testing different models. In Figure 8, we show the changes of PSNR under different noise levels. Among them, our

Table 2. Quantitative results of different single-contrast MRI reconstruction and multi-contrast MRI guided-reconstruction algorithms  $\times 4$  acceleration.

Method	Params	IXI-T2			BrainTS-T2		
		PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$
ZF	-	27.27	0.6188	11.69	28.72	0.6932	9.70
UNet [26]	7.8M	33.24	0.9246	5.85	31.07	0.9443	7.83
MUSC [18]	13.9M	34.39	0.9326	5.23	37.11	0.9644	3.71
Restormer [39]	26.1M	35.79	0.9481	4.40	38.15	0.9699	3.31
CUNet [3]	0.2M	37.30	0.9577	3.61	36.46	0.9629	4.04
MDUNet [36]	5.2M	39.79	0.9715	2.75	35.26	0.9632	5.68
MTrans [8]	86.1M	39.48	0.9700	2.96	34.02	0.9473	5.33
Restormer* [39]	26.1M	41.61	0.9780	2.24	39.91	0.9778	2.75
Ours-S	1.4M	<b>42.18</b>	<b>0.9804</b>	<b>2.10</b>	<b>41.60</b>	<b>0.9821</b>	<b>2.30</b>
Ours-L	5.7M	<b>42.93</b>	<b>0.9821</b>	<b>1.93</b>	<b>42.58</b>	<b>0.9848</b>	<b>2.05</b>

method achieves the best performance, showing better robustness than other methods. In Figure 6, we can see that the noise in the reference image brings serious artifacts to the target image in existing multi-contrast reconstruction methods [3, 36, 8, 39]. By contrast, our method is least affected by noise. Figure 5 (b) show a special case in the BrainTS test set that the left side of the reference image is missing. Other multi-contrast MRI reconstruction methods restore the left side of the target image that has severe artifacts, while our method avoids this problem. It indicates that our decomposition-based model is more robust than other methods when dealing with noise-polluted and inconsistent reference images.

### 5.3. Ablation studies

**Effect of the number of stages.** Figure 7 shows the performance changes under different number of iterations  $T \in [1, 5]$ . From the figure, we can find that the reconstruction performance of the model becomes better with the increase of the number of iterations. Specifically, from  $T = 1$  to  $T = 4$ , the PSNR improves by 3.35dB and 0.72dB on the guided reconstruction and guided SR tasks, respectively. To balance the model performance and the computational com-

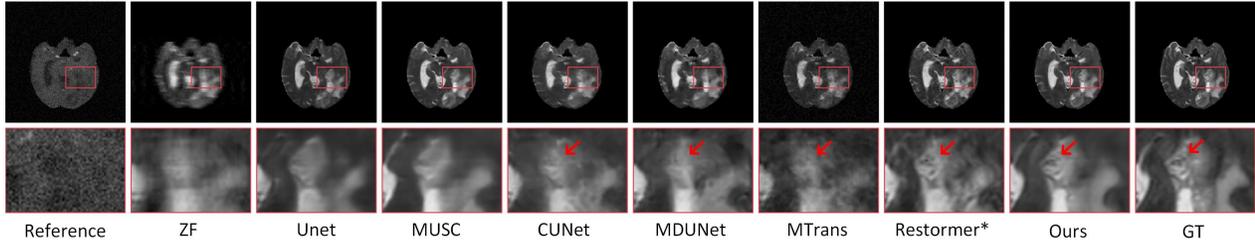


Figure 6. Visual comparison of different models on the BrainTS testing sets with additional noise in reference image.

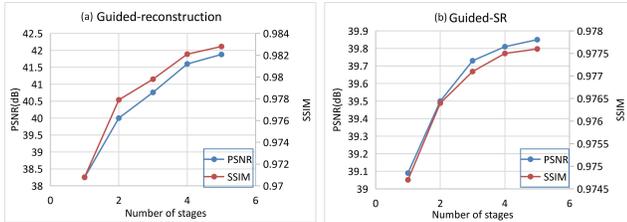


Figure 7. The PSNR and SSIM curves on BrainTS testset with a different number of stages  $T$  for reconstruction and SR tasks.

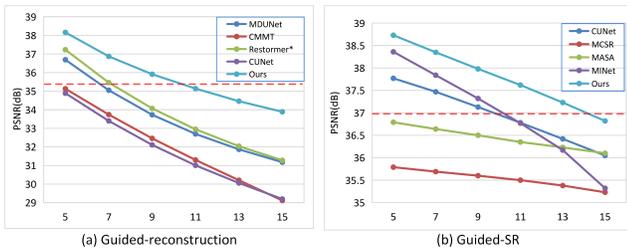


Figure 8. Robustness analysis of additional noise in reference image. The red horizontal lines represent the SOTA results of single-contrast methods.

Table 3. Quantitative comparisons of different MC-VarNet-S configurations on IXI testset for multi-contrast SR task.

<b>Fusion Rule Ablation</b>	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$
No fusion(single contrast)	33.50	0.9292	5.70
Concatenate fusion	41.05	0.9769	2.38
w/o decomposition	41.32	0.9776	2.31
with decomposition(Ours)	<b>41.71</b>	<b>0.9787</b>	<b>2.21</b>
<b>Submodule Ablation</b>	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$
w/o DCL	41.25	0.9774	2.33
with DCL(Ours)	<b>41.71</b>	<b>0.9787</b>	<b>2.21</b>
<b>Denoising Network Ablation</b>	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$
ResNet	41.21	0.9772	2.33
Modified U-Net(Ours)	<b>41.71</b>	<b>0.9787</b>	<b>2.21</b>

plexity, we set  $T = 4$  in our final model.

**Effect of different fusion rules.** We compare the performance of different multi-contrast information fusion rules, including without fusion (single contrast), concatenation fusion, fusion without decomposition, and our decomposition-based fusion. We evaluate these models on multi-contrast MRI SR task. The results of the IXI test set

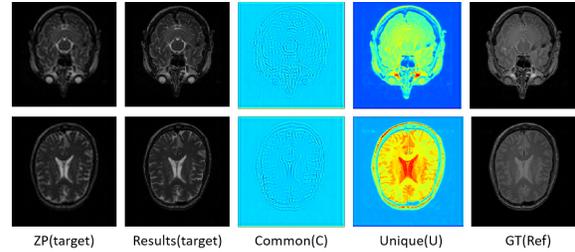


Figure 9. The decomposed common (C) and unique (U) components of the reference image in the last iteration for the multi-contrast SR task.

Table 4. Quantitative comparison of the model with and without channel expansion (CE) operation on BrainTS testset.

Method	SR			Reconstruction		
	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	RMSE $\downarrow$
w/o CE	38.53	0.9726	3.20	39.28	0.9752	2.97
with CE	<b>39.81</b>	<b>0.9775</b>	<b>2.78</b>	<b>41.60</b>	<b>0.9821</b>	<b>2.30</b>

are shown in Table 3. Among them, our method achieves the best performance. In Figure 9, we visualize the decomposed components of our MC-VarNet model on the multi-contrast SR task. As shown in the figure, common components decomposed from the reference image are mainly high-frequency details that are related to the target image. The unique components are mainly the low-frequency information of the reference HR image. It shows the effectiveness of our method in decomposing the reference image into common and unique components.

**Effect of the denoising network and the DCL.** To validate the effectiveness of the modified U-Net for denoising, we replace it with a ResNet [10] denoising module that has a similar number of parameters. As shown in Table 3, the U-Net outperforms the ResNet by 0.5dB PSNR on the IXI test set. DCL ensures the  $k$ -space data consistency between the reconstruction result and the input image. Ablation results in Table 3 show that DCL can effectively improve the model performance by 0.46dB PSNR.

**Effect of the channel expansion operation.** To avoid the problem of information loss caused by channel reduction, we simply repeat the input image along the channel dimension to expand the channel number from 1 to  $c_i$  and

reconstruct  $c_i$  images in the iteration stage simultaneously. The expanded input help the network to learn more diverse information. As listed in Table 4, channel expansion operation improves the performance by 1.28dB PSNR on multi-contrast SR task. On multi-contrast reconstruction task, it also leads to 2.32dB PSNR improvement. The significant performance improvements fully demonstrate the effectiveness of this operation.

## 6. Conclusion

This paper proposed a variational network to solve the multi-contrast MRI SR and reconstruction problem. Different from existing DCNN-based methods that manually designed fusion rules, our MC-VarNet was constructed under the guidance of the optimization algorithm. We designed an optimization algorithm to solve the model and unfolded the iterative solutions into a deep neural network. Thus our model combines the good interpretability of the model-based methods with the powerful feature expression ability of the deep neural networks. Our future works will employ the proposed model to solve more guided-restoration problems, such as RGB-guided depth map SR and guided image denoising.

## 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), the Shanghai Municipal Commission of Economy and Informatization Program (2021-GZL-RGZN-01028) and the East China Normal University Graduate International Conference Special Fund.

## References

- [1] Salman UH Dar, Mahmut Yurt, Mohammad Shahdloo, Muhammed Emrullah Ildiz, Berk Tınaz, and Tolga Çukur. 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.
- [2] Xin Deng and Pier Luigi Dragotti. Deep coupled ista network for multi-modal image super-resolution. *IEEE Transactions on Image Processing*, 29:1683–1698, 2019.
- [3] Xin Deng and Pier Luigi Dragotti. Deep convolutional neural network for multi-modal image restoration and fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3333–3348, 2020.
- [4] Jinming Duan, Jo Schlemper, Chen Qin, Cheng Ouyang, Wenjia Bai, Carlo Biffi, Ghalib Bello, Ben Statton, Declan P O’regan, and Daniel Rueckert. Vs-net: Variable splitting network for accelerated parallel mri reconstruction. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22*, pages 713–722. Springer, 2019.
- [5] Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho-Joon Lee, and Dosik Hwang. Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic resonance in medicine*, 80(5):2188–2201, 2018.
- [6] Chaowei Fang, Dingwen Zhang, Liang Wang, Yulun Zhang, Lechao Cheng, and Junwei Han. Cross-modality high-frequency transformer for MR image super-resolution. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1584–1592, 2022.
- [7] Chun-Mei Feng, Huazhu Fu, Shuhao Yuan, and Yong Xu. Multi-contrast MRI super-resolution via a multi-stage integration network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–149. Springer, 2021.
- [8] Chun-Mei Feng, Yunlu Yan, Geng Chen, Yong Xu, Ying Hu, Ling Shao, and Huazhu Fu. Multi-modal transformer for accelerated MR imaging. *IEEE Transactions on Medical Imaging*, pages 1–1, 2022.
- [9] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [11] Junzhou Huang, Chen Chen, and Leon Axel. Fast multi-contrast MRI reconstruction. *Magnetic resonance imaging*, 32(10):1344–1352, 2014.
- [12] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.
- [13] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.
- [14] Guangyuan Li, Jun Lv, Yapeng Tian, Qi Dou, Chengyan Wang, Chenliang Xu, and Jing Qin. Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast MRI super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20636–20645, 2022.
- [15] Guangyuan Li, Jun Lv, Xiangrong Tong, Chengyan Wang, and Guang Yang. High-resolution pelvic MRI reconstruction using a generative adversarial network with attention and cyclic loss. *IEEE Access*, 9:105951–105964, 2021.
- [16] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1833–1844, 2021.

- [17] Risheng Liu, Shichao Cheng, Long Ma, Xin Fan, and Zhongxuan Luo. Deep proximal unrolling: Algorithmic framework, convergence analysis and applications. *IEEE Transactions on Image Processing*, 28(10):5013–5026, 2019.
- [18] Tianlin Liu, Anadi Chaman, David Belius, and Ivan Dokmanić. Learning multiscale convolutional dictionaries for image reconstruction. *IEEE Transactions on Computational Imaging*, 8:425–437, 2022.
- [19] Liying Lu, Wenbo Li, Xin Tao, Jiangbo Lu, and Jiaya Jia. MASA-SR: Matching acceleration and spatial adaptation for reference-based image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6368–6377, 2021.
- [20] Qing Lyu, Hongming Shan, Cole Steber, Corbin Helis, Chris Whitlow, Michael Chan, and Ge Wang. Multi-contrast super-resolution MRI through a progressive network. *IEEE Transactions on Medical Imaging*, 39(9):2738–2749, 2020.
- [21] Qing Lyu, Hongming Shan, and Ge Wang. MRI super-resolution with ensemble learning and complementary priors. *IEEE Transactions on Computational Imaging*, 6:615–624, 2020.
- [22] Bjoern H. Menze, Andras Jakab, Stefan Bauer, and Kalpathy-Cramer. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.
- [23] Osvald Nitski, Sayan Nag, Chris McIntosh, and Bo Wang. CDF-Net: Cross-domain fusion network for accelerated MRI reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 421–430. Springer, 2020.
- [24] Defu Qiu, Shengxiang Zhang, Ying Liu, Jianqing Zhu, and Lixin Zheng. Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. *Computer methods and programs in biomedicine*, 187:105059, 2020.
- [25] Cong Quan, Jinjie Zhou, Yuanzheng Zhu, Yang Chen, Shanshan Wang, Dong Liang, and Qiegen Liu. Homotopic gradients of generative density priors for mr image reconstruction. *IEEE Transactions on Medical Imaging*, 40(12):3265–3278, 2021.
- [26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [27] Pingfan Song, Lior Weizman, João FC Mota, Yonina C Eldar, and Miguel RD Rodrigues. Coupled dictionary learning for multi-contrast MRI reconstruction. *IEEE Transactions on Medical Imaging*, 39(3):621–633, 2019.
- [28] Pingfan Song, Lior Weizman, João FC Mota, Yonina C Eldar, and Miguel RD Rodrigues. Coupled dictionary learning for multi-contrast MRI reconstruction. *IEEE Transactions on Medical Imaging*, 39(3):621–633, 2019.
- [29] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational networks for accelerated MRI reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 64–73. Springer, 2020.
- [30] Liyan Sun, Zhiwen Fan, Xueyang Fu, Yue Huang, Xinghao Ding, and John Paisley. A deep information sharing network for multi-contrast compressed sensing MRI reconstruction. *IEEE Transactions on Image Processing*, 28(12):6141–6153, 2019.
- [31] Haoqian Wang, Xiaowan Hu, Xiaole Zhao, and Yulun Zhang. Wide weighted attention multi-scale network for accurate MR image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):962–975, 2021.
- [32] Hong Wang, Yuexiang Li, Deyu Meng, and Yefeng Zheng. Adaptive convolutional dictionary network for CT metal artifact reduction. In *The 31st International Joint Conference on Artificial Intelligence*, pages 1401–1407, 2022.
- [33] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3103–3112, June 2020.
- [34] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. In *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 514–517. IEEE, 2016.
- [35] Lior Weizman, Yonina C Eldar, and Dafna Ben Bashat. Reference-based MRI. *Medical physics*, 43(10):5357–5369, 2016.
- [36] Lei Xiang, Yong Chen, Weitang Chang, Yiqiang Zhan, Weili Lin, Qian Wang, and Dinggang Shen. Deep-learning-based multi-modal fusion for fast MR reconstruction. *IEEE Transactions on Biomedical Engineering*, 66(7):2105–2114, 2018.
- [37] Cheng Yang, Risheng Liu, Long Ma, Xin Fan, Haojie Li, and Miao Zhang. Unrolled optimization with deep priors for intrinsic image decomposition. In *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, pages 1–7. IEEE, 2018.
- [38] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. ADMM-CSNet: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(3):521–538, 2018.
- [39] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.
- [40] Jinwei Zhang, Hang Zhang, Chao Li, Pascal Spincemaille, Mert Sabuncu, Thanh D Nguyen, and Yi Wang. Temporal feature fusion with sampling pattern optimization for multi-echo gradient echo acquisition and image reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–242. Springer, 2021.
- [41] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of*

*the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3217–3226, 2020.

- [42] Yulun Zhang, Kai Li, Kunpeng Li, and Yun Fu. MR image super-resolution with squeeze and excitation reasoning attention network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13425–13434, 2021.
- [43] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.
- [44] Xiaole Zhao, Yulun Zhang, Tao Zhang, and Xueming Zou. Channel splitting network for single MR image super-resolution. *IEEE Transactions on Image Processing*, 28(11):5649–5662, 2019.
- [45] Hao Zheng, Faming Fang, and Guixu Zhang. Cascaded dilated dense network with two-step data consistency for MRI reconstruction. *Advances in Neural Information Processing Systems*, 32:1742–1752, 2019.
- [46] Bo Zhou and S Kevin Zhou. Dudornet: learning a dual-domain recurrent network for fast MRI reconstruction with deep t1 prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4273–4282, 2020.