Null Space Matters: Range-Null Decomposition for Consistent Multi-Contrast MRI Reconstruction

Jiacheng Chen*, Jiawei Jiang*, Fei Wu, Jianwei Zheng†

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China
zjw@zjut.edu.cn

Abstract

Consistency and interpretability have long been the critical issues in MRI reconstruction. While interpretability has been dramatically improved with the employment of deep unfolding networks (DUNs), current methods still suffer from inconsistencies and generate inferior anatomical structure. Especially in multi-contrast scenes, different imaging protocols often exacerbate the concerned issue. In this paper, we propose a range-null decomposition-assisted DUN architecture to ensure consistency while still providing desirable interpretability. Given the input decomposed, we argue that the inconsistency could be analytically relieved by feeding solely the null-space component into proximal mapping, while leaving the range-space counterpart fixed. More importantly, a correlation decoupling scheme is further proposed to narrow the information gap for multi-contrast fusion, which dynamically borrows isotropic features from the opponent while maintaining the modality-specific ones. Specifically, the two features are attached to different frequencies and learned individually by the newly designed isotropy encoder and anisotropy encoder. The former strives for the contrast-shared information, while the latter serves to capture the contrast-specific features. The quantitative and qualitative results show that our proposal outperforms most cutting-edge methods by a large margin. Codes will be released on https://github.com/chenjiachengzzz/RNU.

Introduction

As a non-invasive and radiation-free in vivo imaging technique, magnetic resonance imaging (MRI) (Yang et al. 2022; Feng et al. 2021c), has evolved into an indispensable medical resource, delivering exceptional soft tissue contrast and precise anatomical visualization alongside signal quantification. However, the acquisition of high-resolution MR image demands prolonged scan intervals, which poses problems such as patient discomfort and motion-induced distortions. Consequently, a specific investigation emerges to speed up the acquisition process while ensuring a high image quality. The most vital solution lies in an undersampling step in k-space, coupled with a reconstruction algorithm using compressed sensing technique (Song, Chen, and Zhang 2021). This paradigm shift significantly accelerates the MRI acquisition. Unfortunately, due to the extreme violation of Nyquist theorem, the undersampled k-space data often suffer from aliasing artifacts that may spread throughout the image. This phenomenon of inconsistency severely impedes the clinical diagnosis and is surely unacceptable in areas of life and health.

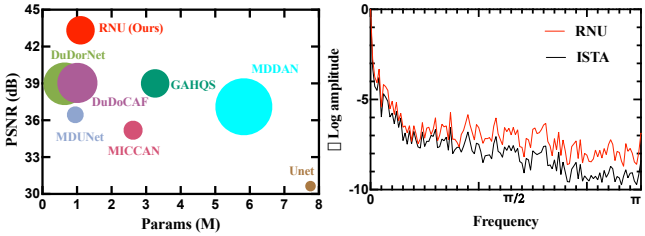

Figure 1: (Left) Comparisons of PSNR/Params-Flops among competing methods. The circle radius denotes the metrics of Flops (G). (Right) Relative log amplitudes of spectral feature maps.

To address the challenges, most conventional approaches borrow certain hand-crafted priors to enforce the reconstructed images enjoy truthful visual properties. For example, techniques such as sparse representation (Wang et al. 2023a), low rankness (Peng et al. 2023), edge sharpness (Huang et al. 2022), and non-local similarity (Wen et al. 2023) have been extensively investigated and yield commendable performance. These solutions incrementally ameliorate image quality through iterative updates, which often suffer from the expense of massive execution time and the risk of oversmoothed recoveries. Moreover, the necessity for meticulous hyperparameter fine-tuning remains a formidable concern. With the contiguous progress of hardware and software, reconstruction methods centering around deep neural networks (DNNs) have emerged as the prevailing paradigm. Early DNN-based reconstruction approaches often borrow architectures from the natural image domain, such as UNet (Ronneberger, Fischer, and Brox 2015), ResNet (He et al. 2016), or Transformer (Vaswani et al. 2017). However, a high-quality performance cannot be guaranteed since these architectures are not designed specifically for MRI reconstruction. Moreover, poor inter-

*These authors contributed equally.
†Corresponding authors.

pretability is also a daunting issue since a purely deep learning network is generally known as a black box. Recently, the field has seen a rapid paradigm shift with the emerging technique of Deep Unfolding Networks (DUNs) (Song, Chen, and Zhang 2023), which take advantage of both traditional algorithms and network modules. On that basis, high performance can be earned without losing much interpretability. Following this line, some pioneers (Feng et al. 2022, 2021a,b; Fang et al. 2022) commence to harness fast-acquired modalities/contrasts, i.e., T1WI and PDWI, to accelerate the reconstruction of counterparts, such as T2WI and FSPDWIs. With the aid of reference information, this multi-contrast MRI acceleration has soon led to substantial enhancements.

Although significant strides have been made, current MRI techniques are not yet accessible to the practical clinical applications. While most DUNs were proposed to strive for better interpretability, their performance cannot be ensured due to an approximated prior setting or brute feature learning. ***To the best of our knowledge, the matter of consistency guarantee, i.e., certain constraints that ensure coherence between reconstruction and ground truth, literally remains unexplored***. Note that similar to the hand-crafted priors, neural networks also endeavor to capture underlying data knowledge and constrain the solution space, which can be considered as an implicit way for reconstruction consistency. Nonetheless, current schemes either rely solely on single-modality information or simply concatenate data from auxiliary/reference modalities, which encounter constraints on feature interaction, culminating in performance degradation. In practice, an arbitrary incorporation of multi-contrast features would bring little information gain for MRI, yet exacerbate the challenge of inter-domain variations. In other words, relying blindly on reference information may negatively introduce inductive biases and engender turbulence in acquisition. Hence, ***a pressing question arises: how to effectively harness the complementary information yet discard the redundant features from opposite contrast?***

For the first issue, we propose to impose a consistency guarantee on typical DUNs. Practically, the range-null decomposition is used to disentangle features into two components. With the range component preserved, we further refine the null space to guarantee the coherence of reconstructed images. The proposal not only attains superior interpretability, but also ensures high credibility for the outcomes. For the second issue, it is widely acknowledged that auxiliary modalities, even originating from the same site within the same patient, exhibit substantial dissimilarities due to the use of different imaging protocols. Henceforth, we expound upon a correlation decoupling learning block (CDLB) that effectively disentangles the overall information into contrast-specific (or anisotropic) part and contrast-shared (or isotropic) one. On one hand, the shared features of both contrasts exhibit similar edges and structures, which could be used to facilitate cross-modality fusion. On the other hand, modality-specific features contain various pathological details and texture information, which should be preserved to promote the reconstruction fidelity. Therefore, we need a divide-and-conquer approach to properly integrate complementary information while retaining exclusive information. Hence, CDLB embodies a dual-branch structure accommodating an isotropy encoder and an anisotropy encoder. More specifically, anisotropy encoder is designed to facilitate the transmission of contrast-specific information, while isotropy encoder is crafted to learn contrast-shared features. On that basis, the reconstruction quality can be enhanced by jointly considering the shared information, which further ensure a high consistency.

The performance of our proposals is validated on a naive unfolding network, i.e., Iterative Shrinkage Thresholding Algorithm (ISTA), without sophisticated structure. In cases with different acceleration factors and sampling patterns, extensive experiments on two representative datasets demonstrate that our method, yet with moderate running complexity, outperforms all previously reported SOTA methods. Several initial comparisons are given in Fig. 1. Evidently, the other competitors lag behind our method by a large margin. Besides, compared to the original ISTA, the addition of consistency guarantee clearly enriches high-frequency information, which favors the sharper texture structures.

Related Work

MRI Acceleration

Magnetic resonance imaging is widely known as a non-invasive medical technique, which has become a critical tool in diagnosing diseases and monitoring treatment outcomes. However, the acquisition of MRI images can be computationally complex, resulting in a prolonged waiting time, which hinders its widespread use in clinical practice. To address the problem, researchers have developed various acceleration techniques, among which compressed sensing and deep learning are the most widely employed. Compressed sensing (Zhou and Li 2023) aims to reduce the amount of data acquired during MRI scans, which in turn improves reconstruction speed. Various compressed sensing techniques, including sparse regularization (Zhang et al. 2022), edge-based sketching (Jiang, Zhai, and Kong 2021), and dictionary learning (Liu et al. 2020), have been proposed. However, these hand-designed methods often require massive iteration steps and suffer poor generalization due to the need for strong prior information (Hammernik et al. 2018). Recently, deep learning methods have also shown promising results in promoting MRI reconstruction speeds. Convolutional neural networks (CNNs) are the pivotal branch to reconstruct MRI images from highly undersampled k-space data, achieving high-quality results at fast speeds. Schlemper et al. (Chlemper et al. 2017) develop a deep cascade of CNN for dynamic MR image reconstruction, which outperforms compressed sensing approaches in terms of both speed and reconstruction error. Similarly, MICCAN (Huang et al. 2019) recovers high-quality MR details from a cascaded channel-wise attention that can attend to prominent features while filtering out irrelevant information. Unlike these methods, Wang et al. (Wang et al. 2020) design a network that performs directly on the complex values, namely DeepcomplexMRI, which is able to preserve the intrinsic relationship between the real and imaginary parts. Then,

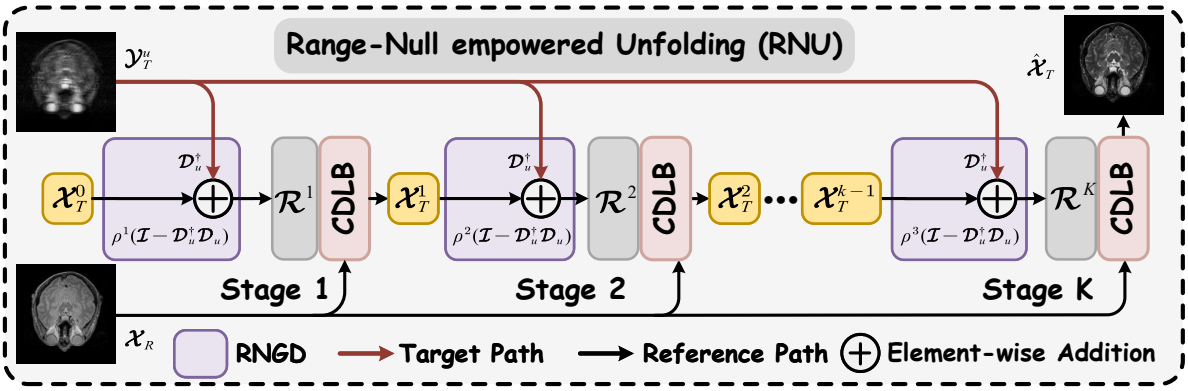

Figure 2: Illustration of our Range-Null empowered Unfolding network. Each stage is composed of a range-null empowered data fidelity term and a correlation decoupling learning module.

Feng et al. (Feng et al. 2021d) introduce OctConv(Chen et al. 2019) to address the shortcoming that CNN cannot compute complex numbers. Since artifacts in the image domain are often non-local and structural, some researchers commence to introduce hybrid domain learning paradigm that exploits both image space and k-space. Zhou et al. (Zhou and Zhou 2020) propose a dual domain recurrent network (DuDoRNet) to simultaneously recover k-space data and images. It is experimentally proven to outperform single domain methods. Other approaches enforce the use of unsupervised learning and generative adversarial networks (GANs)(Huang et al. 2022; Jiang et al. 2021) to reduce the reliance on labeled data. Nevertheless, models based on GAN tend to yield fake details, which is unallowable in medical domain.

Deep Unfolding Network

Overall, the aforementioned techniques have shown promising results in accelerating MRI reconstruction and hold the potential to improve clinical practice. However, they either lack interpretability or exhibit unstable reconstruction quality. Hence, the combination of compressed sensing theory and deep learning techniques has gained in-depth investigation to improve the accuracy and speed of MRI reconstruction. Accordingly, deep unfolding network has opened a new avenue in this field, which mainly follows the optimization steps of certain traditional algorithms, yet with partial components replaced with network modules. The learned version of iterative soft thresholding algorithm (Zheng et al. 2023) (ISTA) is known as the first DUN, whose theoretical and experimental outcomes evince that the specific configuration not only enjoys an adaptive update of network parameters in an end-to-end fashion, but also effectively preserves the interpretability of the iterative optimization. Subsequently, an array of DUNs have surfaced. As an example, a fresh approach founded on half-quadratic splitting (HQS) has been formulated to expedite the reconstruction of undersampled MR images (Jiawei et al. 2023). Furthermore, MGDUN (Yang et al. 2022) contributes significantly to the field of image super-resolution, which introduces an end-to-end trainable model that harnesses the power of multiple contrasts. Additionally, PLA-GAN (Jiang et al. 2023) presents a pioneering ADMM technique for CT denoising, which merges the flexibility of model-based methodologies with the benefits of generative adversarial networks.

Proposed Methodology

Problem Formulation

Fast MRI reconstruction is considered to recover the target image $\boldsymbol{\mathcal{X}}_T \in \mathbb{R}^{H\times W}$ from the k-space undersampled measurement $\boldsymbol{\mathcal{Y}}_T^u \in \mathbb{C}^{H\times W}$. The typical optimization framework can be formulated as follows.

$$\hat{\boldsymbol{\mathcal{X}}}_T = \arg\min_{\boldsymbol{\mathcal{X}}_T} \underbrace{\|\boldsymbol{\mathcal{Y}}_T^u - \boldsymbol{\mathcal{D}}_u\boldsymbol{\mathcal{X}}_T\|_F^2}_{\text{data fidelity}} + \lambda \underbrace{\boldsymbol{\mathcal{P}}(\boldsymbol{\mathcal{X}}_T)}_{\text{prior}} \tag{1}$$

where λ is a trade-off parameter used to balance the inter-term contributions, $\boldsymbol{\mathcal{X}}_T$ is the recovered result, and $\boldsymbol{\mathcal{D}}_u$ denotes the undersampling operator in fast MRI. Besides, $\boldsymbol{\mathcal{P}}(\cdot)$ denotes the regularization term.

Despite the advancements made in single-contrast MRI, employing multiple measurements would probably facilitate a better understanding of the imaging process. As a result, much of the attention on MRI acceleration has been drawn toward the multi-contrast reconstruction. Accordingly, the Eq. (1) can be rewritten as

$$\hat{\boldsymbol{\mathcal{X}}}_T = \arg\min_{\boldsymbol{\mathcal{X}}_T} \|\boldsymbol{\mathcal{Y}}_T^u - \boldsymbol{\mathcal{D}}_u\boldsymbol{\mathcal{X}}_T\|_F^2 + \lambda\boldsymbol{\mathcal{P}}(\boldsymbol{\mathcal{X}}_T, \boldsymbol{\mathcal{X}}_R) \tag{2}$$

where $\boldsymbol{\mathcal{X}}_R$ denotes a reference image that is typically full-sampled. Since different modalities would reflect altered characteristics of the same patient, the utilization of complementary information potentially benefits a high-quality reconstruction of the target images.

Following the typical ISTA framework, Eq. (2) can be solved by iterating between the following gradient descent and proximal mapping steps:

$$\boldsymbol{\mathcal{R}}^k = \boldsymbol{\mathcal{X}}_T^{k-1} - \rho\boldsymbol{\mathcal{D}}_u^T(\boldsymbol{\mathcal{D}}_u\boldsymbol{\mathcal{X}}_T^{k-1} - \boldsymbol{\mathcal{Y}}_T^u) \tag{3a}$$

$$\boldsymbol{\mathcal{X}}_T^k = \boldsymbol{prox}_{\boldsymbol{\mathcal{P}},\lambda}(\boldsymbol{\mathcal{R}}^k, \boldsymbol{\mathcal{X}}_R) \tag{3b}$$

where $\boldsymbol{\mathcal{R}}^k$ and $\boldsymbol{\mathcal{X}}_T^k$ denote the intermediate result and reconstruction MR image at k-th stage, respectively. ρ is the

step size of gradient descent. $\boldsymbol{\mathcal{D}}_u^T$ denotes the transpose of $\boldsymbol{\mathcal{D}}_u$. $\boldsymbol{prox}_{\boldsymbol{\mathcal{P}},\lambda}(\boldsymbol{v}) = \arg\min_{\boldsymbol{v}} \frac{1}{2}\|\boldsymbol{u}-\boldsymbol{v}\|_2^2 + \lambda\boldsymbol{\mathcal{P}}(\boldsymbol{u})$. The limitation of Eq. (3b) is that gradient descent can only find an approximate solution to the data fidelity term in each iteration, which cannot strictly satisfy $\boldsymbol{\mathcal{Y}}_T^u = \boldsymbol{\mathcal{D}}_u\boldsymbol{\mathcal{R}}$. Note again that when dealing with medical data concerning closely to healthcare, any trivial mistake may lead to serious consequences. ***Hence, ensuring reconstruction consistency is an urgent need.***

Range-Null Empowered Unfolding (RNU) Network

To tackle the bottleneck of ISTA-based DUN, we introduce the Range-Null Space decomposition to strive for the consistency guarantee of the data fidelity term. Given an undersampled matrix $\boldsymbol{\mathcal{D}}_u$, at least one pseudo-inverse $\boldsymbol{\mathcal{D}}_u^\dagger$ that satisfies $\boldsymbol{\mathcal{D}}_u\boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{D}}_u = \boldsymbol{\mathcal{D}}_u$ exists (Wang et al. 2023b). Hence, any MR image $\boldsymbol{\mathcal{X}}_T$ can be decomposed into two components:

$$\boldsymbol{\mathcal{X}}_T = \underbrace{\boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{D}}_u}_{\text{range-space}}\boldsymbol{\mathcal{X}}_T + \underbrace{(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{D}}_u)}_{\text{null-space}}\boldsymbol{\mathcal{X}}_T \tag{4}$$

where operator $\boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{D}}_u$ projects $\boldsymbol{\mathcal{X}}_T$ from the original domain into range space, and $\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{D}}_u$ projects MR image to null space. Generally, MRI reconstruction can be separately treated within the two distinct spaces. Range space reconstruction ensures the data consistency with respect to the observation $\boldsymbol{\mathcal{Y}}_T^u$, while null space refinement tends to remove artifacts and promote image quality. More importantly, $\boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{D}}_u\boldsymbol{\mathcal{X}}_T = \boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{Y}}_T^u$ can be directly achieved from $\boldsymbol{\mathcal{Y}}_T^u$, hence only the null-space shall be further ameliorated. Due to the relatively limited information in null space, the model burden as well as the convergence capacity would be naturally boosted. Hence, Eq. (3) can be rewritten as:

$$\boldsymbol{\mathcal{R}}^k = \boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{Y}}_T^u + \rho^k(\boldsymbol{\mathcal{I}} - \boldsymbol{\mathcal{D}}_u^\dagger\boldsymbol{\mathcal{D}}_u)\boldsymbol{\mathcal{X}}_T^{k-1} \tag{5a}$$

$$\boldsymbol{\mathcal{X}}_T^k = \boldsymbol{prox}_{\boldsymbol{\mathcal{P}},\lambda}(\boldsymbol{\mathcal{R}}^k, \boldsymbol{\mathcal{X}}_R) \tag{5b}$$

Note the role of the learnable parameter ρ^k has now been shifted into dynamically balance the contribution of range space and null space.

As shown in Fig. 2, our network architecture is stage-wisely armed with two main components, i.e., **Range-Null based Gradient Descent** (RNGD) and **Correlation Decoupling Learning Block** (CDLB). By feeding $\boldsymbol{\mathcal{Y}}_T^u$ and $\boldsymbol{\mathcal{X}}_T^{k-1}$ into RNGD module, $\boldsymbol{\mathcal{R}}^k$ can be easily updated using Eq. (5a). With the support of theoretically strict Range-Null decomposition, the data consistency is readily enriched, which further guarantees a high reconstruction quality.

Correlation Decoupling Learning Block Given $\boldsymbol{\mathcal{R}}^k$ updated, CDLB is designed to cope with Eq. (5b) and generate the reconstruction result $\boldsymbol{\mathcal{X}}_T^k$. Three modules are involved, including an encoder for feature extraction, a decoder for image generation, and a decoupling and fusion segment for information aggregation. As illustrated in Fig. 3, all the three modules follow a dual-branch architecture. To facilitate a better consistency of the reconstituted images, the main endeavor is paid towards assembling a wealth of complementary information, while discarding the modality-specific details. Note that multi-modality MRI data are acquired through distinct imaging protocols from identical anatomical sites of the same patient. Hence, it is natural that, while the local details show a certain anisotropy, the global characterization bears an isotropy with a large probability. Moreover, Fig. 4 visualizes some representative multi-contrast features with frequency divided into high- and low-parts. As can be observed, the low-frequency features exhibit congruence in their patterns, yet the high-frequency features embody a discrepancy. Motivated by these findings, we respectively refer low-frequency long-range information and high-frequency local details as the *isotropic* features and *anisotropic* features. During the reconstruction, we prioritize the isotropic features to be fused, which borrows the more correlated information from the auxiliry contrast and discard the less concerned ones.

Encoder. The main role of Encoder block is to extract features $\{\boldsymbol{\mathcal{F}}_T^E, \boldsymbol{\mathcal{F}}_R^E\}$ from target and reference inputs:

$$\boldsymbol{\mathcal{F}}_T^E = \boldsymbol{\mathcal{E}}_T(\boldsymbol{\mathcal{H}}_T^{IN}(\boldsymbol{\mathcal{X}}_T^k)), \boldsymbol{\mathcal{F}}_R^E = \boldsymbol{\mathcal{E}}_R(\boldsymbol{\mathcal{H}}_R^{IN}(\boldsymbol{\mathcal{X}}_R)) \tag{6}$$

where $\boldsymbol{\mathcal{H}}_T^{IN}(\cdot)$ and $\boldsymbol{\mathcal{H}}_R^{IN}(\cdot)$ denote the input projection composing of a 3×3 convolutional layer followed a ReLU activation function, respectively for target and reference input. $\boldsymbol{\mathcal{E}}(\cdot)$ is formed by multiple isotropy encoders, whose details will be elaborated shortly.

Isotropy Encoder. To extract low-frequency isotropy features, most pre-defined operators can be readily used, such as Discrete Wavelet Transform and Discrete Fourier Transform. However, these static operations are not flexible enough to earn for compelling performance. As mentioned, isotropy features refer to the global characterization of the multi-contrast data. On that basis, we dynamically capture the low-frequency structure using transformer architecture, which enjoys a high reputation in recognizing long-range dependencies. Concretely, given half the features $\{\boldsymbol{\mathcal{F}}_T^{E,I}, \boldsymbol{\mathcal{F}}_R^{E,I}\}$ split from $\{\boldsymbol{\mathcal{F}}_T^E, \boldsymbol{\mathcal{F}}_R^E\}$, isotropy encoder $\boldsymbol{\mathcal{I}}(\cdot)$ can be mathematically expressed as:

$$\boldsymbol{\mathcal{F}}_T^I = \boldsymbol{\mathcal{I}}_T(\boldsymbol{\mathcal{F}}_T^{E,I}), \boldsymbol{\mathcal{F}}_R^I = \boldsymbol{\mathcal{I}}_R(\boldsymbol{\mathcal{F}}_R^{E,I}) \tag{7}$$

where $\boldsymbol{\mathcal{F}}_T^I$ and $\boldsymbol{\mathcal{F}}_R^I$ are the isotropy feature of target and reference MRI, respectively. Note that we not only employ $\boldsymbol{\mathcal{I}}(\cdot)$ for the purpose of frequency division, but also embed it in the overall encoder block due to its function in capturing global features.

The original transformer suffers from $\mathcal{O}(n^2)$ complexity. To boost the efficiency yet with the guarantee of exploring globally the anatomical structure, our isotropy encoder improves the typical transformer architecture with two new modules, as shown in Fig. 3(C). Taking $\boldsymbol{\mathcal{F}}_T^{E,I}$ as an instance, $\boldsymbol{\mathcal{I}}(\cdot)$ can be further formulated as:

$$\boldsymbol{\mathcal{F}}_{FMM} = \boldsymbol{\mathcal{F}}_T^{E,I} + \boldsymbol{\mathcal{H}}_{FMM}(\boldsymbol{\mathcal{H}}_{LN}(\boldsymbol{\mathcal{F}}_T^{E,I})) \tag{8}$$

$$\boldsymbol{\mathcal{F}}^I = \boldsymbol{\mathcal{F}}_{FMM} + \boldsymbol{\mathcal{H}}_{LKFFN}(\boldsymbol{\mathcal{H}}_{LN}(\boldsymbol{\mathcal{F}}_{FMM})) \tag{9}$$

where $\boldsymbol{\mathcal{H}}_{FMM}(\cdot)$ and $\boldsymbol{\mathcal{H}}_{LKFFN}(\cdot)$ denote the Fourier modulation module (FMM) and large-kernel based feed-forward network (LKFFN), respectively. $\boldsymbol{\mathcal{H}}_{LN}(\cdot)$ refers to the layer normalization operation.

FMM is used to replace the high-consumptional self-attention computation. Commonly, given any normalized

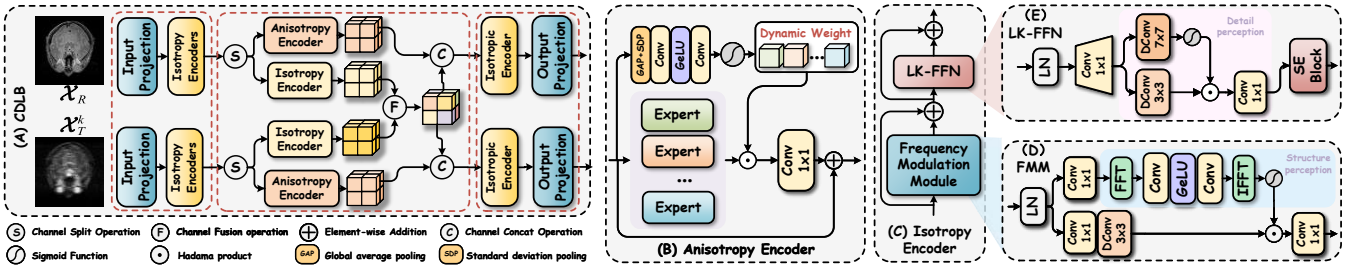

Figure 3: (A) Illustration of Correlation Decoupling Learning Block. (B) Anisotropy Encoder. (C) Isotropy Encoder. (D) Frequency Modulation Module. (E) Large Kernel based FFN.

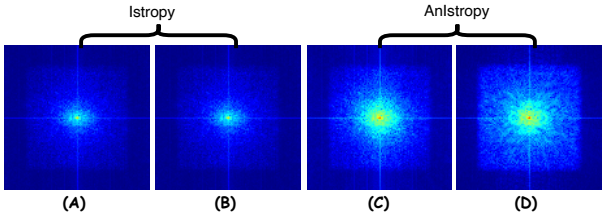

Figure 4: Fourier spectrum visualization of multi-contrast data. (A)-(B) Low-frequency parts. (C)-(D) High-frequency parts.

tensor $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{H\times W\times C}$, we shall obtain $\boldsymbol{\mathcal{Q}}$ and $\boldsymbol{\mathcal{V}}$ with some convolution operations, e.g., convolution with 1×1 kernel,

$$\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{H}}_{1\times1}(\boldsymbol{\mathcal{F}}), \boldsymbol{\mathcal{V}} = \boldsymbol{\mathcal{H}}_{1\times1}(\boldsymbol{\mathcal{F}}) \tag{10}$$

However, convolutions capture only the local features. Based on the Fast Fourier Transformation (FFT), the images in frequency domain are the weighting of signals in image domain. Thus, we can learn global structure simply with convolutions in frequency domain.

$$\boldsymbol{\mathcal{A}}_{fft} = \sigma(\psi^{-1}(\boldsymbol{\mathcal{H}}_{1\times1}(\delta(\boldsymbol{\mathcal{H}}_{1\times1}(\psi(\boldsymbol{\mathcal{Q}})))))) \tag{11}$$

where $\boldsymbol{\mathcal{A}}_{fft}$ represents the global attention map. ψ and ψ^{-1} denote the FFT and inverse FFT, respectively. σ and δ denote sigmoid operation and GeLU function, respectively. Finally, we produce the enriched features by:

$$\boldsymbol{\mathcal{F}}_{FMM} = \boldsymbol{\mathcal{F}} + \boldsymbol{\mathcal{H}}_{1\times1}(\boldsymbol{\mathcal{A}}_{fft}\boldsymbol{\mathcal{V}}) \tag{12}$$

LKFFN is specified to improve the naive feed-forward network (FFN), in which we introduce a gated mechanism based on recent advances in large kernel convolution, discriminatively enhancing the perception of local details. Given a layer normalized feature $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{H\times W\times C}$, LKFFN can be formulated as:

$$\hat{\boldsymbol{\mathcal{F}}} = \boldsymbol{\mathcal{H}}_{SE}(\boldsymbol{\mathcal{H}}_{1\times1}(\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{F}}))) + \boldsymbol{\mathcal{F}} \tag{13}$$

$$\boldsymbol{\mathcal{G}}(\boldsymbol{\mathcal{F}}) = \delta(\boldsymbol{\mathcal{H}}_{7\times7}(\boldsymbol{\mathcal{H}}_{1\times1}(\boldsymbol{\mathcal{F}}))) \odot \boldsymbol{\mathcal{H}}_{3\times3}(\boldsymbol{\mathcal{H}}_{1\times1}(\boldsymbol{\mathcal{F}})) \tag{14}$$

where $\odot$ denotes element-wise multiplication, $\boldsymbol{\mathcal{H}}_{SE}(\cdot)$ is employed to mix the channel information. Generally, gating function $\boldsymbol{\mathcal{G}}(\cdot)$ allows the network to focus on texture details, which are fitly supplementary to the role of FMM that learns Fourier attention maps.

Anisotropy Encoder. Several conventional operators, such as Sobel and Canny, are readily employable to extract high-frequency anisotropy, yet the hand-crafted priors may deviate far from the information flow during dynamic update. As noted, for multi-contrast data, anistropic features refer mainly to local details. Hence, we consider CNN-based treatments to be naturally suitable for highlighting the pathological subtleties contained in high-frequency information, due to their known adeptness in integrating short-range dependencies. Similar to $\boldsymbol{\mathcal{I}}(\cdot)$, given halved features $\{\boldsymbol{\mathcal{F}}_T^{E,A}, \boldsymbol{\mathcal{F}}_R^{E,A}\}$ from $\{\boldsymbol{\mathcal{F}}_T^{E}, \boldsymbol{\mathcal{F}}_R^{E}\}$, anistropy encoder $\boldsymbol{\mathcal{A}}(\cdot)$ is formulated as:

$$\boldsymbol{\mathcal{F}}_T^{A} = \boldsymbol{\mathcal{A}}_T(\boldsymbol{\mathcal{F}}_T^{E,A}), \boldsymbol{\mathcal{F}}_R^{A} = \boldsymbol{\mathcal{A}}_R(\boldsymbol{\mathcal{F}}_R^{E,A}) \tag{15}$$

where $\boldsymbol{\mathcal{F}}_T^{A}$ and $\boldsymbol{\mathcal{F}}_R^{A}$ are anisotropic features of target and reference, respectively. Considering that both the edges and textures of organs are vital for medical analysis, $\boldsymbol{\mathcal{A}}(\cdot)$ is expected to perceptualize as comprehensive information as possible. Therefore, we choose various convolution operations to parallelly behave as experts for detecting different features, involving standard convolution with 3×3 and 5×5 kernels, depth-wise convolution with 7×7 kernel, and dilated convolution with 5×5 and 7×7 kernels. In addition, to measure the individual contribution of each expert, a self-attention mechanism is employed to learn the concerned weights. As shown in Fig. (3) (B), given an input $\boldsymbol{\mathcal{F}}^{E,A}$, we first apply global average and standard deviation poolings to generate channel descriptor $\boldsymbol{\mathcal{Z}}_a$ and $\boldsymbol{\mathcal{Z}}_s$:

$$\boldsymbol{\mathcal{Z}}_{a,c} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{\mathcal{F}}_c^{E,A}(i,j) \tag{16}$$

$$\boldsymbol{\mathcal{Z}}_{s,c} = \sqrt{\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(\boldsymbol{\mathcal{F}}_c^{E,A}(i,j) - \boldsymbol{\mathcal{Z}}_{a,c})} \tag{17}$$

where (i,j) and c denote position and channel, respectively. Then a gating mechanism is used to aggregate channel-wise information, from which the weight $\boldsymbol{\mathcal{W}}$ can be allocated:

where W_U and W_D refer to different weights from two 1×1 convolutional layers. Afterwards, the expert-aggregated feature can be dynamically learned as:

$$\boldsymbol{\mathcal{F}}^{A} = \boldsymbol{\mathcal{H}}_{1\times1}([\boldsymbol{\mathcal{W}}[m]\cdot\boldsymbol{\mathcal{F}}_{exp}^{m}]), m = 1,2,...,M \tag{18}$$

where $\boldsymbol{\mathcal{F}}_{exp}^{m}$ and M represent the learned expert feature and the overall expert number, respectively. $[\cdot]$ denotes the channel-wise concatenation.

Fusion Layer. With the isotropic and anisotropic characteristics individually extracted, the role of the fusion layer is to borrow the isotropy information from reference to facilitate target reconstruction. In practice, to maintain the consistency of inductive bias, we again apply $\boldsymbol{\mathcal{I}}(\cdot)$ for the isotropy aggregation, that is:

$$\boldsymbol{\mathcal{F}}_F^I = \boldsymbol{\mathcal{I}}(\boldsymbol{\mathcal{F}}_T^I, \boldsymbol{\mathcal{F}}_R^I) \tag{19}$$

where $\boldsymbol{\mathcal{F}}_F^I$ denotes the fused feature of target and reference modalities, which is further fed into decoder branch.

Decoder. In the decoder stage, the separated isotropy and anisotropy features are channel-wisely concatenated to regenerate the final MR images, which is formulated as:

$$\boldsymbol{\mathcal{F}}_T = \boldsymbol{\mathcal{H}}_T^O(\boldsymbol{\mathcal{D}}_T([\boldsymbol{\mathcal{F}}_F^I, \boldsymbol{\mathcal{F}}_T^A])), \boldsymbol{\mathcal{F}}_R = \boldsymbol{\mathcal{H}}_R^O(\boldsymbol{\mathcal{D}}_R([\boldsymbol{\mathcal{F}}_F^I, \boldsymbol{\mathcal{F}}_R^A])) \tag{20}$$

where $\boldsymbol{\mathcal{F}}_T$ and $\boldsymbol{\mathcal{F}}_R$ denote the results of target and reference contrasts, respectively. Note again that for a consistent semantic flow, our decoder follows a symmetrical structure with encoder. That is to say, $\boldsymbol{\mathcal{H}}^O(\cdot)$ and $\boldsymbol{\mathcal{D}}(\cdot)$ share the same operations with $\boldsymbol{\mathcal{H}}^{IN}(\cdot)$ and $\boldsymbol{\mathcal{E}}(\cdot)$.

Experiments

Datasets and Implementation Details

Datasets: The classical IXI dataset and the currently largest fastMRI (Zbontar et al. 2018) dataset are employed for performance evaluation. IXI consists of 578 paired PDWI and T2WI volumes. FastMRI is paired with multi-contrast images. Following (Feng et al. 2022), we also sift out 240 pairs of PDWI and FS-PDWI knee images in our experiments.

Metrics: Two popular metrics, including peak signal-to-noise ratio (PSNR) and structure similarity (SSIM), are used to measure the quality of reconstruction results.

Implementation Details: The proposed RNU is implemented using PyTorch and evaluated with NVIDIA 3090 GPU. The Adam optimizer is utilized for model training, with an initial learning rate $lr = 10^{-4}$ that is gradually decayed to 10^{-6} over 50 epochs. The batch size is set as 4. To facilitate a better generalization, the training data are randomly augmented by flipping horizontally or vertically and rotating at different angles. L_1 loss is used to optimize the network. To ensure a fair comparison, all competing approaches are trained using the finely tuned parameter settings. Unless specified otherwise, the stage number K is 8.

Results and Analysis

Quantitative Results At $4\times$ or $8\times$ acceleration rate and under three masks, i.e., random, radial, and Gaussian sampling, the achieved results of all competing methods on both datasets are presented in Table 1. Note all values are averaged from 5 run for stability and fairness. Evidently, irrespective of sampling masks, acceleration rates, and varying datasets, RNU consistently stands out as a new SOTA method. In terms of the PSNR results on IXI with $\times 4$ random mask, the leads of RNU over Unet (Ronneberger, Fischer, and Brox 2015), MICCAN (Huang et al. 2019), MDUNet (Xiang et al. 2018), DuDorNet (Zhou and Zhou 2020), MDDAN (Yang et al. 2020), DuDoCAF(Lyu et al. 2022), and GAHQS (Jiawei et al. 2023) reach 12.69, 8.14, 6.88, 4.36, 6.23, 4.29, and 4.33, respectively. These results substantiate the advantages of our proposed consistency guarantee using range-null decomposition.

Qualitative Results The comparisons of target contrast and the concerned reconstruction error are given in Figs. 5 and 6. The texture of error maps serves as an indicator of restoration efficacy, whereby a smoother and bluer surface signifies a better reconstruction. As evidenced, given the input suffering substantial aliasing artifacts and lacking anatomical details, RNU is capable of reconstructing the more details and exhibiting the fewer visible artifacts, surpassing all the other competitors significantly. The enhancement in quality can again be attributed to the employment of consistency guarantee both in data fidelity and proximal mapping. Moreover, the comprehensive exploitation of supplementary features from multi-contrast data further contributes to the refinement of the ultimate outcomes.

Ablation Experiments

Effect of Unfolding Stages To analyze the evolution of performance improvements with increasing iterations, the results of our model adapting to different stages, that is, from 4 to 12, are presented in Fig. 7. Remarkably, the increasing stages consistently promote the model performance to higher levels, albeit at the expense of a heavier computational load. As evidenced in Table 1, our RNU outperforms other competitors significantly by setting $K = 8$. This lead endows us with an easy-to-control model with respect to practical hardware resources and performance requirements.

Effect of Rang-Null Decomposition The comparisons of RNU and original ISTA are also given in Fig. 7. Note that, for fairness, the proximal mapping step in ISTA follows the same network architecture with our RNU. The only difference lies in the employment of rang-null decomposition. Clearly, RNU outperforms the counterpart in all stages. The large margin directly substantiates our conjecture that coducting fidelity update in the whole space would lead to suboptimal data consistency. Moreover, the null space holds more potential for information correction and noise removal.

Effect of FMM We compare FMM with alternative self-attention mechanisms and elucidate the findings in Table 2. To ensure a fair comparison, the model parameters using different attentions are standardized to a similar level of 1.1 million. As can be seen, local attention (Wang et al. 2022) achieves linear complexity with significant performance degradation. In other words, the efficiency boost of local attention is at the cost of losing the global modeling capabilities. With the aid of fast Fourier transformation, the meticulously designed FMM not only achieves outcomes akin to those of global attention (Vaswani et al. 2017) , but also ensures the linear model complexity.

Effect of LKFFN The introduction of LKFFN endeavors to enlarge the receptive field, yet with the guarantee of limited extra computational demands. As evidenced in Table 3, with only 0.2M, 0.15M, and 0.02M Params added, the gains of our model over FFN (Vaswani et al. 2017), ConvFFN

AR	Methods	IXI						fastMRI					
		Random		Radial		Gaussian		Random		Radial		Gaussian	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×4	Unet	30.636	0.947	33.634	0.963	33.591	0.964	27.792	0.808	28.619	0.829	25.586	0.737
	MICCAN	35.187	0.972	42.776	0.993	43.008	0.992	29.112	0.835	30.186	0.865	28.800	0.825
	MDUNet	36.444	0.982	37.314	0.983	36.984	0.982	28.835	0.824	29.366	0.840	26.407	0.758
	DuDorNet	38.967	0.984	44.855	0.994	45.844	0.995	29.034	0.834	30.004	0.861	26.990	0.825
	MDDAN	37.097	0.983	40.666	0.990	42.583	0.992	29.192	0.837	30.081	0.862	blue28.649	0.840
	DuDoCAF	39.037	0.985	45.001	0.994	45.636	0.994	29.163	0.837	30.244	0.866	27.723	0.808
	GAHQS	38.997	0.984	43.666	0.992	44.583	0.993	29.134	0.835	30.014	0.860	27.890	0.827
	RNECD (Ours)	**43.323**	**0.994**	**50.374**	**0.999**	**49.661**	**0.998**	**29.949**	**0.851**	**31.228**	**0.878**	**30.320**	**0.873**
×8	Unet	28.662	0.927	29.379	0.929	30.879	0.943	26.330	0.752	26.256	0.720	23.634	0.662
	MICCAN	31.554	0.949	33.919	0.967	36.058	0.977	27.377	0.769	27.297	0.737	26.282	0.730
	MDUNet	35.176	0.973	34.163	0.976	35.444	0.979	27.451	0.773	26.906	0.737	25.182	0.700
	DuDorNet	36.193	0.974	38.272	0.982	41.106	0.990	27.367	0.774	27.137	0.742	25.643	0.767
	MDDAN	35.383	0.979	35.611	0.979	37.425	0.983	27.659	0.781	27.307	0.748	26.697	0.768
	DuDoCAF	36.319	0.976	37.574	0.977	40.587	0.987	27.557	0.776	27.515	0.751	27.723	0.808
	GAHQS	35.977	0.973	37.163	0.976	40.444	0.984	27.467	0.774	27.117	0.712	25.653	0.767
	RNECD (Ours)	**40.894**	**0.991**	**40.901**	**0.991**	**41.927**	**0.993**	**28.379**	**0.795**	**28.047**	**0.766**	**28.335**	**0.797**

Table 1: The numerical results under three masks as well as × 4 and × 8 acceleration rates. The best and second results are highlighted and underlined, respectively.

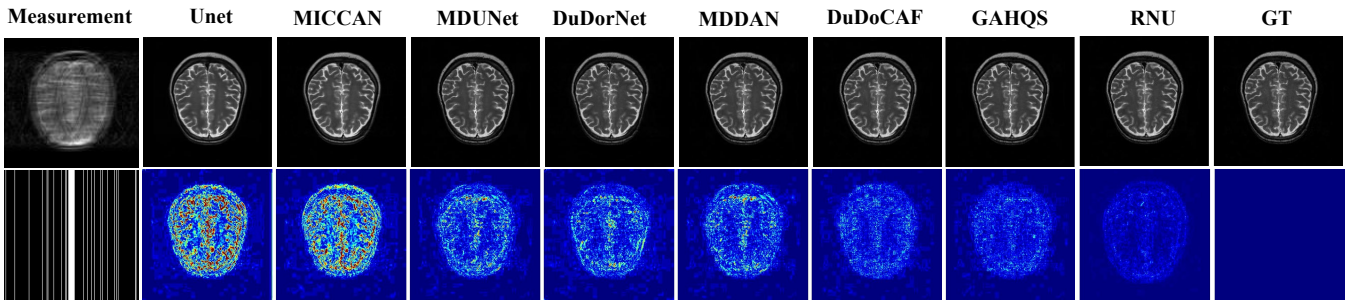

Figure 5: Visual results and error maps on IXI dataset with random mask and ×8 acceleration rate.

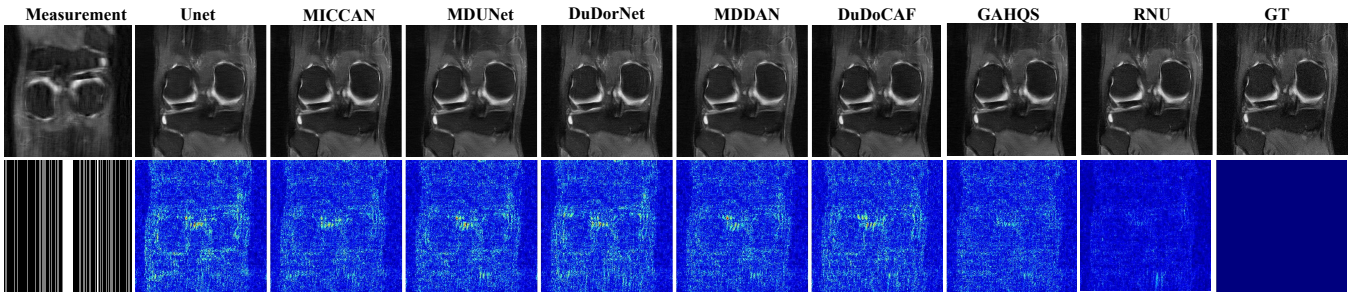

Figure 6: Visual results and error maps on fastMRI dataset with random mask and ×4 acceleration rate.

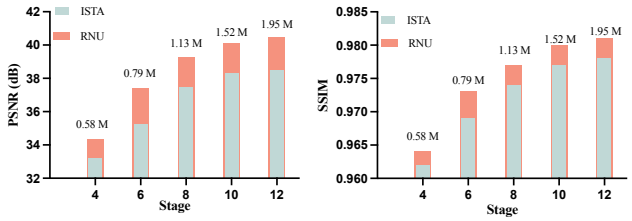

Figure 7: Comparison on different number of stages on IXI dataset with gaussian mask and $\times 4$ acceleration rate.

(Wang et al. 2022), and GDFN (Zamir et al. 2022) reach 0.595dB, 0.441dB, and 0.246dB, respectively. These leads are notably evident, since all the competitors have been reported recently as the state-of-the-art schemes.

Method	Complexity	PSNR	SSIM
Global MSA	$\mathcal{O}(N^2)$	38.199	0.973
Local W-MSA	$\mathcal{O}(N)$	37.595	0.968
FMM	$\mathcal{O}(N)$	**38.287**	**0.977**

Table 2: Ablation of different attention mechanisms on IXI dataset under $\times 4$ random acceleration mask.

Method	FFN	ConvFFN	GDFN	LKFFN
PSNR(db)	38.579	38.733	38.928	**39.174**
SSIM	0.982	0.983	0.986	**0.988**
Params (M)	0.93	0.98	1.11	1.13

Table 3: Ablation of Feed-Forward Network on IXI dataset under $\times 4$ random acceleration mask.

Effect of CDLB To substantiate the favorable impact of CDLB, a series of ablation studies involving diverse multi-contrast fusion schemes have been conducted. As shown in Table 4, our proposal achieves notable gains of 3.976 dB and 1.427 dB in terms of PSNR, respectively, over single-modality and traditional multi-modality treatments. The superiority can be attributed to the fundamental information interplay facilitated by CDLB, which enables the amalgamation of shared features that encompass high-sharpness intricacies, while dampening the modality-specific information within the low-frequency domain. Furthermore, as shown in Fig. 8, our reconstructed image truly enjoys better consistency with GT. Through spectral analysis, it can be observed that the CDLB equipped results hold much trivial error to the true value. Recall that in Fig. 4, the spectrograms in different frequencies are achieved specifically by the Isotropy and Anisotropy encoders. As anticipated, our prior supposition holds true: Isotropy encoder adeptly acquires low-frequency details, while Anisotropy encoder excels in assimilating high-frequency characteristics.

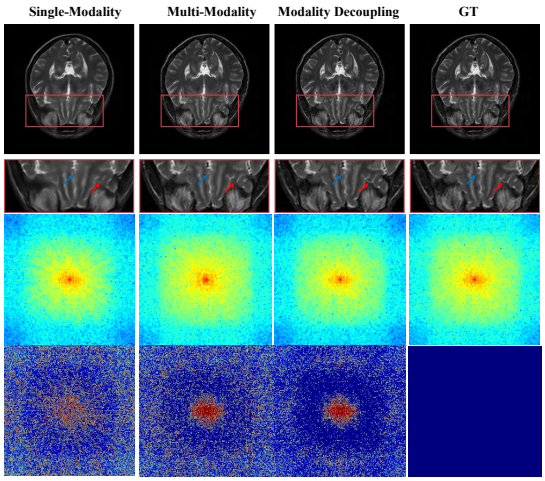

Figure 8: Visual results from different modality treatments on IXI with random mask and $\times 8$ acceleration rate. The first and second rows show the reconstruction images and localized enlargements, respectively. The third and fourth rows show the spectrograms and error plots in Fourier domain, respectively.

Method	PSNR	SSIM
Single-Modality	35.198	0.975
Multi-Modality	37.747	0.985
Modality decoupling	**39.174**	**0.988**

Table 4: Ablation of multi-contrast fusion on IXI dataset under $\times 4$ random acceleration mask.

Conclusion

In this paper, we propose a novel Range-Null decomposition empowered Unfolding network (RNU) for fast multi-contrast MR imaging. To ensure a better consistency, RNU commences to analytically analyze the inertial property of range component, then bestows greater emphasis on null space to strive for coherence in the reconstructed MRIs. Moreover, to strengthen the isotropy yet suppress the anisotropy of multiple contrasts, we proceed to formulate a scheme for correlation decoupling that not only harmonizes supplementary information but also maintains the integrity of the inherent features. Ultimately, by seamlessly integrating the dual mechanisms into the conventional unfolding architecture, the overall RNU model emerges. Comprehensive MRI acceleration experiments on two representative datasets, also with masks and acceleration rates varied, show that our method are superior to the previous leading approaches by a large margin. We hope this study will provide valuable insights for the community to design efficient and effective MRI reconstruction architectures.

Acknowledgments

This work was supported in part by the Key Program of Natural Science Foundation of Zhejiang Province, China, under

Grant LZ24F030012, and in part by the National Natural Science Foundation of China under Grant 62276232.

References

Chen, Y.; Fan, H.; Xu, B.; Yan, Z.; Kalantidis, Y.; Rohrbach, M.; Yan, S.; and Feng, J. 2019. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 3435–3444.

Chlemper, J.; Caballero, J.; Hajnal, J.; Price, A.; and Rueckert, D. 2017. A deep cascade of convolutional neural networks for dynamic MR image reconstructio. *IEEE Transactions on Medical Imaging*, 37(2): 491–503.

Fang, C.; Zhang, D.; Wang, L.; Zhang, Y.; Cheng, L.; and Han, J. 2022. Cross-modality high-frequency transformer for MR image super-resolution. In *Proceedings of the 30th ACM International Conference on Multimedia (ACMMM)*, 1584–1592.

Feng, C.-M.; Fu, H.; Yuan, S.; and Xu, Y. 2021a. Multi-contrast mri super-resolution via a multi-stage integration network. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, 140–149. Springer.

Feng, C.-M.; Yan, Y.; Chen, G.; Xu, Y.; Hu, Y.; Shao, L.; and Fu, H. 2022. Multi-modal transformer for accelerated mr imaging. *IEEE Transactions on Medical Imaging*.

Feng, C.-M.; Yan, Y.; Yu, K.; Xu, Y.; Shao, L.; and Fu, H. 2021b. Exploring separable attention for multi-contrast MR image super-resolution. *arXiv preprint arXiv:2109.01664*.

Feng, C.-M.; Yang, Z.; Chen, G.; Xu, Y.; and Shao, L. 2021c. Dual-octave convolution for accelerated parallel MR image reconstruction. In *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 116–124.

Feng, C.-M.; Yang, Z.; Fu, H.; Xu, Y.; Yang, J.; and Shao, L. 2021d. DONet: dual-octave network for fast MR image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13. Doi:10.1109/TNNLS.2021.3090303.

Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M. P.; Sodickson, D. K.; Pock, T.; and Knoll, F. 2018. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6): 3055–3071.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Huang, J.; Ding, W.; Lv, J.; Yang, J.; Dong, H.; Del Ser, J.; Xia, J.; Ren, T.; Wong, S. T.; and Yang, G. 2022. Edge-enhanced dual discriminator generative adversarial network for fast MRI with parallel imaging using multi-view information. *Applied Intelligence*, 52(13): 14693–14710.

Huang, Q.; Yang, D.; Wu, P.; Qu, H.; Yi, J.; and Metaxas, D. 2019. MRI reconstruction via cascaded channel-wise attention network. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, 1622–1626.

Jiang, J.; Feng, Y.; Xu, H.; and Zheng, J. 2023. Low-dose CT reconstruction via optimization-inspired GAN. In *In 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5.

Jiang, M.; Zhai, F.; and Kong, J. 2021. A novel deep learning model DDU-net using edge features to enhance brain tumor segmentation on MR images. *Artificial Intelligence in Medicine*, 121: 102180.

Jiang, M.; Zhi, M.; Wei, L.; Yang, X.; Zhang, J.; Li, Y.; Wang, P.; Huang, J.; and Yang, G. 2021. FA-GAN: Fused attentive generative adversarial networks for MRI image super-resolution. *Computerized Medical Imaging and Graphics*, 92: 101969.

Jiawei, J.; Jiacheng, C.; Honghui, X.; Yuchao, F.; and Jianwei, Z. 2023. GA-HQS: MRI reconstruction via a generically accelerated unfolding approach. In *International Conference on Multimedia and Expo (ICME)*, 1–6.

Liu, Q.; Yang, Q.; Cheng, H.; Wang, S.; Zhang, M.; and Liang, D. 2020. Highly undersampled magnetic resonance imaging reconstruction using autoencoding priors. *Magnetic Resonance in Medicine*, 83(1): 322–336.

Lyu, J.; Sui, B.; Wang, C.; Tian, Y.; Dou, Q.; and Qin, J. 2022. DuDoCAF: Dual-domain cross-attention fusion with recurrent transformer for fast multi-contrast MR imaging. In *International Conference on Medical Image Computing and ComputerAssisted Intervention (MICCAI)*, 474–484.

Peng, J.; Wang, Y.; Zhang, H.; Wang, J.; and Meng, D. 2023. Exact decomposition of joint low rankness and local smoothness plus sparse matrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5766–5781.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and ComputerAssisted Intervention (MICCAI)*, 234–241.

Song, J.; Chen, B.; and Zhang, J. 2021. Memory-augmented deep unfolding network for compressive sensing. In *Proceedings of the 29th ACM International Conference on Multimedia (ACMMM)*, 4249–4258.

Song, J.; Chen, B.; and Zhang, J. 2023. Deep memory-augmented proximal unrolling network for compressive sensing. *International Journal of Computer Vision*, 131(1): 1477–1496.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems (NIPS)*, 30.

Wang, J.; Ding, D.; Li, Z.; Feng, X.; Cao, C.; and Ma, Z. 2023a. Sparse tensor-based multiscale representation for point cloud geometry compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 9055–9071.

Wang, S.; Cheng, H.; Ying, L.; Xiao, T.; Ke, Z.; Zheng, H.; and Liang, D. 2020. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magnetic Resonance Imaging*, 68: 136–147.

Wang, Y.; Hu, Y.; Yu, J.; and Zhang, J. 2023b. Gan prior based null-space learning for consistent super-resolution. In *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2724–2732.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17683–17693.

Wen, Y.-H.; Gao, L.; Fu, H.; Zhang, F.-L.; Xia, S.; and Liu, Y.-J. 2023. Motif-GCNs with local and non-Local temporal blocks for skeleton-based action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2009–2023.

Xiang, L.; Chen, Y.; Chang, W.; Zhan, Y.; Lin, W.; Wang, Q.; and Shen, D. 2018. Deep-learning-based multi-modal fusion for fast MR reconstruction. *IEEE Transactions on Biomedical Engineering*, 66(7): 2105–2114.

Yang, G.; Zhang, L.; Zhou, M.; Liu, A.; Chen, X.; Xiong, Z.; and Wu, F. 2022. Model-guided multi-contrast deep unfolding network for MRI super-resolution reconstruction. In *Proceedings of the 30th ACM International Conference on Multimedia (ACMMM)*, 3974–3982.

Yang, Y.; Wang, N.; Yang, H.; Sun, J.; and Xu, Z. 2020. Model-driven deep attention network for ultra-fast compressive sensing MRI guided by cross-contrast MR image. In *International Conference on Medical Image Computing and ComputerAssisted Intervention (MICCAI)*, 188–198.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5728–5739.

Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M. J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*.

Zhang, M.; Zhang, M.; Zhang, F.; Chaddad, A.; and Evans, A. 2022. Robust brain MR image compressive sensing via re-weighted total variation and sparse regression. *Magnetic Resonance Imaging*, 85: 271–286.

Zheng, Z.; Dai, W.; Xue, D.; Li, C.; Zou, J.; and Xiong, H. 2023. Hybrid ISTA: Unfolding ISTA with convergence guarantees using free-form deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3226–3244.

Zhou, B.; and Zhou, S. K. 2020. DuDoRNet: learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4273–4282.

Zhou, S.; and Li, G. Y. 2023. Federated learning via inexact ADMM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 9699–9708.