DuDoCAF: Dual-Domain Cross-Attention Fusion with Recurrent Transformer for Fast Multi-contrast MR Imaging

Jun Lyu[1], Bin Sui[1], Chengyan Wang[2(✉)], Yapeng Tian[3], Qi Dou[4], and Jing Qin[5]

[1] School of Computer and Control Engineering, Yantai University
[2] Human Phenome Institute, Fudan University
wangcy@fudan.edu.cn
[3] University of Rochester
[4] Dept. of Computer Science and Engineering, The Chinese University of Hong Kong
[5] Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University

Abstract. Multi-contrast magnetic resonance imaging (MC-MRI) has been widely used for the diagnosis and characterization of tumors and lesions, as multi-contrast MR images are capable of providing complementary information for more comprehensive diagnosis and evaluation. However, it usually suffers from long scanning time to acquire multi-contrast MR images; in addition, long scanning time may lead to motion artifacts, degrading the image quality. Recently, many studies have proposed to employ the fully-sampled image of one contrast with short acquisition time to guide the reconstruction of the other contrast with long acquisition time so as to speed up the scanning. However, these studies still have two shortcomings. First, they simply concatenate the features of the two contrast images together without digging and leveraging the inherent and deep correlation between them. Second, as aliasing artifacts are complicated and non-local, sole image domain reconstruction with local dependencies is far from enough to eliminate these artifacts and achieve faithful reconstruction results. We present a novel **Du**al-**Do**main **C**ross-**A**ttention **F**usion (DuDoCAF) scheme with recurrent transformer to comprehensively address these shortcomings. Specifically, the proposed CAF scheme enables deep and effective fusion of features extracted from two modalities. The dual-domain recurrent learning allows our model to restore signals in both k-space and image domains, and hence more comprehensively remove the artifacts. In addition, we tame recurrent transformers to capture long-range dependencies from the fused feature maps to further enhance reconstruction performance. Extensive experiments on public fastMRI and clinical brain datasets demonstrate that the proposed DuDoCAF outperforms the state-of-the-art methods under different under-sampling patterns and acceleration rates.

Keywords: MRI Reconstruction · Cross-attention Fusion · Recurrent Transformer · Dual-domain Reconstruction

1 Introduction

Magnetic resonance imaging (MRI) is widely used in clinical practice, as it is non-invasive and capable of providing superior soft tissue contrast. Multi-contrast (MC) MR images are obtained from different pulse sequences, which form the image intensity changes between different tissues [14,11]. For example, in brain examination, the T1 weighted images (T1WIs) are used for observing the morphological information, while the fluid attenuated inversion recovery (FLAIR) images are used to detect the edema and inflammation[12]. Similarly, in knee imaging, proton density weighted images (PDWIs) provide the knee structure information while fat-suppressed proton density weighted images (FS-PDWIs) can suppress fat signals and highlight cartilage ligaments[2]. Unfortunately, MR imaging is inherently time-consuming as data are acquired sequentially in k-space. The total scanning time for a typical clinical protocol is about 15$\sim$20 minutes. To this end, reconstructing high-quality MC images from limited acquired measurements to reduce scanning time is highly demanded in practice.

Recently, several studies[17,15,20,3,4,11,21,5] demonstrated that it is a promising way to employ a fully-sampled reference image of one contrast with short acquisition time, such as T1WI and PDWI, to reconstruct the under-sampled target image of the other contrast with longer scanning time, such as FLAIR and FS-PDWI. A key concern of this reconstruction task is that how to fuse the two or more MC images so that their complementary information can be sufficiently leveraged. Early studies either simply stack the MC images in the input layer [17,15] or extract MC features in different branches separately and then stack information in deeper layers, which ignore the inherent and rich correlations among different contrasts. Later, generative adversarial network (GAN) based models, such as rsGAN[3] and Y-net[4], have been developed for synergistic recovery of under-sampled multi-contrast acquisitions. Some multi-scale integration networks [11,5] have also been proposed to extract multi-scale information among the contrasts and incorporate data consistency units for MRI acceleration. Although certain improvement has been made, these algorithms still lack of effective mechanisms to sufficiently harness the correlations between MC images. In addition, the long-range dependencies in the fused features are not well modeled for more faithful reconstruction.

Recently, many transformer based models [7,10] have been introduced to capture global interactions between contexts for fast MRI reconstruction. Feng et al.[7] proposed a task transformer for multi-task learning, which allows to transfer shared structure representation to the task specific branch for MRI reconstruction and super-resolution. Korkmaz et al.[10] introduced zero-shot learned adversarial transformers for unsupervised reconstruction in accelerated MRI. However, these methods focus on restoring images in mono-space and do not exploit any k-space information, which is essential for our task. As each signal in the k-space is estimated from all the values in image domain via Fourier transform, only using constrains in sole image-space cannot effectively reconstruct high-quality aliasing-free images. In this regard, a dual domain recurrent network (DuDoRNet)[21] is proposed to accelerate MR imaging, which demonstrates the

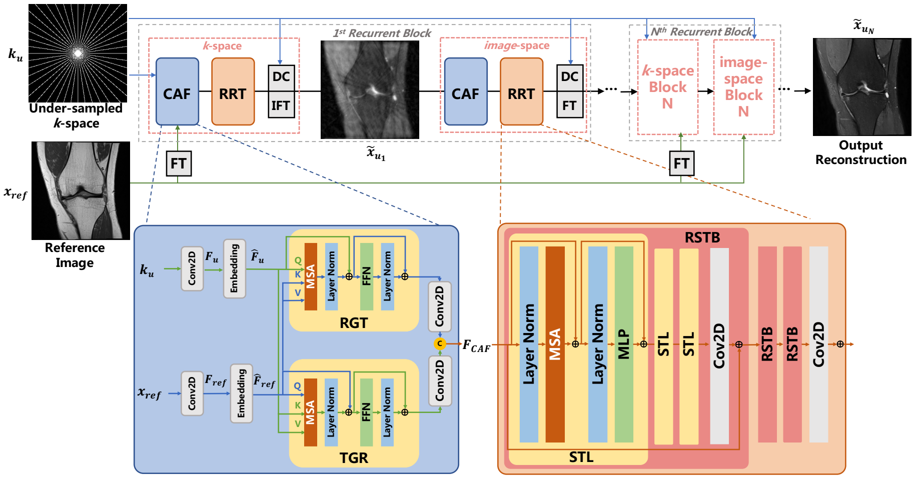

Fig. 1: The architecture of our proposed Dual-Domain Cross-Attention Fusion based Recurrent Transformer (DuDoCAF) for fast multi-contrast MR Imaging.

advantages of cross domain learning. Inspired by this work, we propose a novel MC MR reconstruction method via cross domain learning.

To address above-mentioned limitations, we propose a novel dual-domain cross-attention fusion mechanism (DuDoCAF) with recurrent transformer for fast multi-contrast MR Imaging. Unlike existing models that merely concatenate the features of MC MR images, the proposed CAF mechanism is able to deeply and effectively fuse the features extracted from these two contrast images in a bidirectional way so that complementary information of two contrasts can be sufficiently harnessed. We further tame the residual-reconstruction transformer to model the long-range dependencies based on the fused feature maps in both domains to counteract aliasing artifacts and faithfully reconstruct the target images. In addition, the recurrent dual-domain learning makes the reconstruction results more interpretable, which is important in clinical practice. Extensive experiments on two representative datasets demonstrate that our proposed method gains remarkable margins over several state-of-the-art methods under different sampling patterns and acceleration factors.

2 Methods

2.1 Network Architecture

The network architecture of the proposed model is illustrated in Fig. 1. We denote the complex-valued fully-sampled k-space as k. The corresponding image x reconstructed from k can be obtained by $x = \mathcal{F}^{-1}k$, where $\mathcal{F}^{-1}$ is the inverse Fourier Transform (IFT). To accelerate MR imaging, we employ binary mask M to define the under-sampling trajectory, *e.g.* cartesian, radial, and spiral. Thus, the under-sampled k-space can be defined as $k_u = Mk$, and correspondingly, $x_u = \mathcal{F}^{-1}Mk$. Given the under-sampled k-space data k_u of the target modality (e.g., FLAIR and FS-PDWI), and the fully-sampled image x_{ref} of the reference modality (e.g., T1WI and PDWI), we aim to reconstruct the MR image $\tilde{x}_{u_N}$ from k_u, where N is the number of recurrent blocks in the proposed model. As shown in Fig.1, the proposed model is mainly composed of three modules: 1) the cross-attention fusion (CAF) module, which is employed to deeply and effectively fuse information of different modalities, 2) the residual-reconstruction transformer (RRT), which is harnessed to more faithfully reconstruct the target modality in both domains by capturing more long-range dependencies in the fused k-space and image, 3) the recurrent restoration blocks with data consistency (DC) layer in both k-space and image-space. We spell out their mechanisms as follows.

2.2 Cross-Attention Fusion

In order to ensure that the reference image can effectively guide the target image reconstruction, we need to fuse the two different contrast images. Inspired by [1,18,13], we designed the cross-attention fusion module to establish a bidirectional correspondence between the reference and target images and perform dual feature aggregation.

First, to extract shallow features of under-sampled target $k_u \in \mathbb{R}^{H\times W\times 2}$ and the reference $k_{ref} \in \mathbb{R}^{H\times W\times 2}$, we employ a 3×3 convolutional layer conv to obtain an initial representation $\mathbf{F}_u = \mathrm{Conv}(k_u)$, $\mathbf{F}_{ref} = \mathrm{Conv}(k_{ref}) \in \mathbb{R}^{H\times W\times C}$. Then, the features are reshaped into non-overlapping local windows of size $N\times N$ with the number $\frac{H\times W}{N^2}$. We obtain $\hat{\mathbf{F}}_u, \hat{\mathbf{F}}_{ref} \in \mathbb{R}^{d\times C}$ after the embedding operation, where $d = \frac{H\times W}{N^2}N^2$. The CAF block consists of two sub-modules: reference guide target (RGT) and target guide reference (TGR), and its mechanism can be formulated as:

$$\mathbf{F}_{CAF} = \mathrm{concat}(\mathrm{conv}(\mathrm{RGT}(\hat{\mathbf{F}}_u, \hat{\mathbf{F}}_{ref}), \mathrm{conv}(\mathrm{TGR}(\hat{\mathbf{F}}_u, \hat{\mathbf{F}}_{ref}))), \tag{1}$$

where $\mathbf{F}_{CAF}$ indicates the fused feature of the two contrast images, which will be fed into the following RRT module. The operator concat means the channel-wise concatenation operator.

Then, the attention mechanism jointly learns the $W_Q^{ref}, W_K^{ref}, W_V^{ref}$ and W_Q^u, W_K^u, W_V^u, which are the query (Q), key (K), and value (V) weight matrices for reference and target images, respectively. Here, all weight metrics have the same dimensions $d\times d$. Take the RGT for example, the attention is calculated by encoding target as queries and taking reference as keys and values:

$$\begin{gathered} Q_u = \hat{\mathbf{F}}_u W_Q^{ref}, K_{ref} = \hat{\mathbf{F}}_{ref} W_K^{ref}, V_{ref} = \hat{\mathbf{F}}_{ref} W_V^{ref}, \\ \mathrm{Attention} = \mathrm{softmax}\left(\frac{Q_u \cdot K_{ref}^T}{\sqrt{d}}\right) V_{ref}. \end{gathered} \tag{2}$$

The TGR is similar to RGT, except that the reference is encoded to queries and target is used as keys and values and thus form a cross-attention mechanism.

As described in [16], the attention can be learned over multiple heads in parallel. If the attention is split into H heads, the dimension of the output of each head is $d_{head} = \frac{d}{H}$. Then, the multi-head self-attention (MSA) mechanism is implemented to extract information from different representation subspaces:

$$\begin{gathered}\text{MultiHeadAttn} = \text{concat}\left(\text{head}_1, \cdots, \text{head}_{\text{H}}\right) W^O, \\ \text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right),\end{gathered} \tag{3}$$

where $W_i^Q \in \mathbb{R}^{d \times d_{head}}$, $W_i^K \in \mathbb{R}^{d \times d_{head}}$, $W_i^V \in \mathbb{R}^{d \times d_{head}}$ and $W^O \in \mathbb{R}^{d \times d}$ are weights to be learned. Next, the output is sent to a feed-forward Network (FFN) block consisted of two linear transformation with ReLU activation, defined as:

$$\text{FFN}(x) = \max\left(0, xW_1 + b_1\right) W_2 + b_2, \tag{4}$$

where W_1,W_2 and b_1,b_2 are the weight matrices and biases vectors, respectively. The LayerNorm (LN) layer is added before both MSA and FFN, and the residual connection is employed for both parts.

2.3 Residual Reconstruction Transformer

The long-range dependencies embedded in the fused feature maps obtained from the CAF are essential for efficient and robust image reconstruction, as it is depending on these dependencies that we reconstruct the target image based on under-sampled signals. To effectively capture these long-range dependencies, we develop the RRT module, which consists of three residual swin transformer block (RSTB) and a convolutional layer. It can be formulated as:

$$\begin{gathered}F_i = H_{RSTB_i}\left(F_{i-1}\right), i = 1, 2, 3, \\ F_{RRT} = H_{conv}\left(F_i\right),\end{gathered} \tag{5}$$

where $F_0 = F_{CAF}$. Each RSTB contains a patch embedding operator, three cascaded swin transformer layers (STL) [11], a patch unembedding operator, a convolution, and a residual connection between the input and output of RSTB. It can be expressed as:

$$\begin{aligned} &F_{i,0} = \mathrm{H}_{\mathrm{Emb}_i}\left(F_{i-1}\right) \\ &F_{i,j} = \mathrm{H}_{\mathrm{STL}_{i,j}}\left(F_{i,j-1}\right), \quad j = 1, 2, 3 \\ &F_i = \mathrm{H}_{\mathrm{CONV}_i}\left(\mathrm{H}_{\mathrm{Unemb}_i}\left(F_{i,j}\right) + F_{i-1}\right)\end{aligned} \tag{6}$$

where $\mathrm{H}_{\mathrm{Emb}_i}(\cdot)$ is the patch embedding from $F_{i-1} \in \mathbb{R}^{H \times W \times C}$ to $F_{i,0} \in \mathbb{R}^{HW \times C}$, and $\mathrm{H}_{\mathrm{Unemb}\ i}(\cdot)$ is the patch unembedding from $F_{i,j} \in \mathbb{R}^{HW \times C}$ to $\mathbb{R}^{H \times W \times C}$. The whole process of the STL can be expressed as:

$$\begin{aligned} &F' = \mathrm{H}_{\mathrm{(S)W-MSA}}\left(\mathrm{H}_{\mathrm{LN}}(F)\right) + F \\ &F'' = \mathrm{H}_{\mathrm{MLP}}\left(\mathrm{H}_{\mathrm{LN}}\left(F'\right)\right) + F'\end{aligned} \tag{7}$$

where F and F'' are the input and output of the STL. $\mathrm{H}_{\mathrm{MLP}}(\cdot)$ and $\mathrm{H}_{\mathrm{LN}}(\cdot)$ denote the multilayer perceptron and the layer normalization layer. Windows multi-head self-attention (W-MSA) and shifted windows multi-head self-attention (SW-MSA) $\mathrm{H}_{\mathrm{(S)W-MSA}}(\cdot)$ are alternatively applied in consecutive STLs.

2.4 Dual-domain Recurrent Learning

Each recurrent block of DuDoCAF contains a k-space block, an image reconstruction block, and two interleaved data consistency (DC) layers. In i-th restoration block of image domain, the optimization can be expressed as minimizing the following model:

$$\arg\min_{\theta_{img}}(\|M\mathcal{F}\mathcal{H}_{img}(x_{u_i}, x_{ref}; \theta_{img}) - k_u\|_2^2 + \lambda\left\|x_f - \mathcal{H}_{img}\left(x_{u_i}, x_{ref}; \theta_{img}\right)\right\|_2^2), \tag{8}$$

where $\mathcal{H}_{img}$ is the image restoration network with parameters θ_{img}, the input of the network is x_{u_i} which comes from the $(i-1)$-th k-space reconstruction block and x_{ref} is the fully-sampled reference image. The first term is the data consistency constraint that ensures the consistency of the reconstruction image in k-space and the second term is a regularization term models the relationship between the reconstructed image $\tilde{x}$ and the fully-sampled image x_f. Similarly, the k-space reconstruction optimization can be formulated as:

$$\arg\min_{\theta_k}(\|M\mathcal{H}_k(k_{u_i}, k_{ref}; \theta_k) - k_u\|_2^2 + \lambda\left\|k_f - \mathcal{H}_k\left(k_{u_i}, k_{ref}; \theta_k\right)\right\|_2^2), \tag{9}$$

where $\mathcal{H}_k$ is the k-space reconstruction network with parameters θ_k. Therefore, the final loss for DuDoCAF is $\mathcal{L} = \sum_{i=1}^{N}\left(\mathcal{L}_{img_i} + \mathcal{L}_{k_i}\right)$, in which N represents the number of recurrent blocks.

3 Experiments

Datasets and Implementation We evaluate our proposed method on two raw MRI datasets: **1)** Clinical brain MRI dataset, which was collected using a 3T Philips Ingenia MRI system (Philips Healthcare, Best, the Netherlands) scanner, including T1W and FLAIR imaging. The dataset consists of 36 healthy subjects and 5 patients. We randomly selected 616 images for training and 250 images (150 from healthy subject and 100 from patients) for testing. **2)** Public fastMRI[9] dataset with paired multi-contrast DICOM images. Following[19,6], we filtered out 240 pairs of PDWI and FS-PDWI knee images, 400 for training and the rest 80 images for testing. Both datasets are real k-space data and the matrix size are 256×256. The under-sampling masks include 8× random, 10× radial and 10× spiral pattern.

Experiments were carried out on a system equipped with GPUs of NVIDIA Tesla V100 (4 cores, each with 32 GB memory). The Adam optimizer [8] is used for the training. The model used a batch size of 2 and learning rate of 10^{-4} for 200 epochs. We set $N = 3$ recurrent groups in our network. For fair comparison, the competing methods all use their default parameter settings. Code will be available at `https://github.com/XAIMI-Lab/DuDoCAF`.

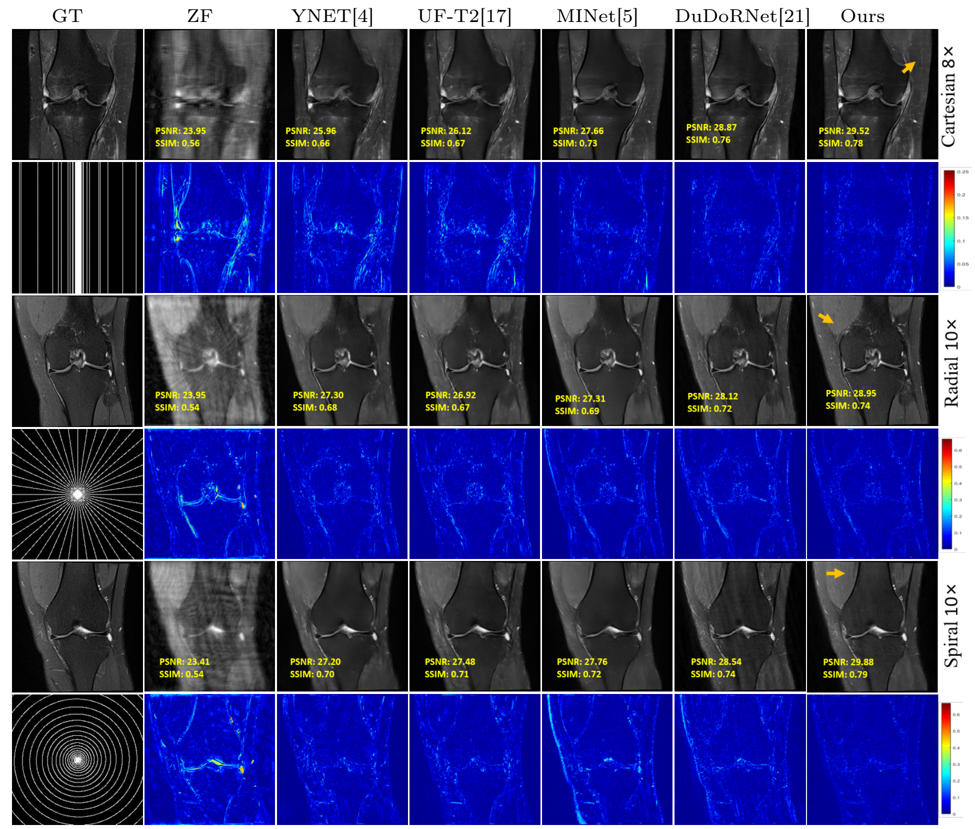

Fig. 2: Reconstruction results from different under-sampling trajectory. The sampling pattern mask and difference images are shown on the second, fourth, and sixth row.

Comparison with state-of-the-arts On both datasets, we have compared our approach with four recent state-of-the-art methods including: YNet[4], UF-T2[17], MINet[5] and DuDoRNet[21].The calculated FLOPs (G) and Parameters (M) of all mentioned models are listed in the supplementary material. Fig.2 shows that knee images reconstructed using YNet and UF-T2 still have aliasing artifacts that are obvious at the edge and in the vessel area. The MINet, DuDoRNet and Ours greatly improved the ZF image quality by recovering sharpness and adding more structural details to the ZF images. However, the yellow arrow shows that, as for fine vessel information, Ours has better reconstruction performance. The reconstructed brain images are shown in supplementary material.

Fig. 3 shows the intermediate results of DuDoCAF with 10× radial undersampling. We can observe the gradual improvement of the reconstruction quality from iteration block N1 to N3. Table 1 shows the mean ± std PSNR and SSIM

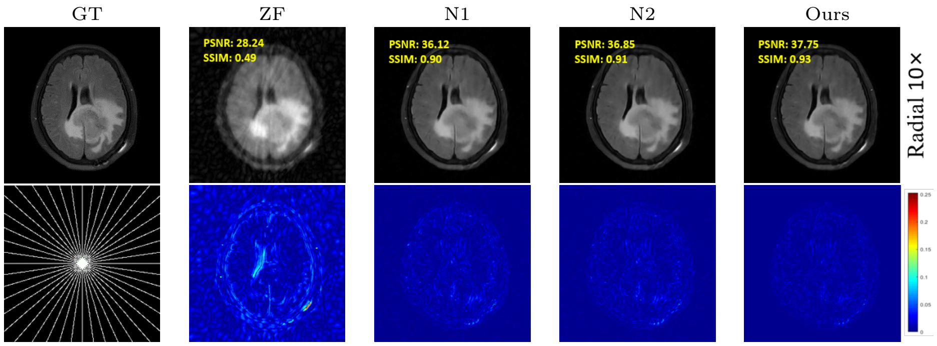

Fig. 3: Visual comparison of DuDoCAF results from each recurrent block using radial sampling (10×).

Table 1: Quantitative results on two datasets with different under-sampling masks, in terms of SSIM and PSNR. The best and second-best results are marked in red and blue, respectively.

Dataset	Methods	Random (8×)		Radial (10×)		Spiral (10×)	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
fastMRI	Y-net	24.96±1.77	0.79±0.01	25.58±1.62	0.81±0.04	25.13±1.51	0.78±0.04
	UF-T2	24.99±1.88	0.79±0.01	25.81±1.55	0.81±0.04	25.80±1.63	0.79±0.04
	MINet	25.32±2.03	0.81±0.01	26.19±1.94	0.82±0.03	26.11±2.28	0.81±0.03
	DuDoRNet	26.14±1.72	0.83±0.01	28.01±1.52	0.86±0.03	26.76±1.91	0.83±0.04
	Ours	27.45±1.52	0.86±0.01	28.91±1.32	0.88±0.03	28.90±1.59	0.87±0.03
Brain	Y-net	33.91±2.71	0.92±0.02	35.20±2.49	0.31±0.03	33.41±3.90	0.95±0.03
	UF-T2	34.02±2.62	0.93±0.02	35.97±2.57	0.95±0.02	33.95±3.95	0.95±0.03
	MINet	36.32±2.77	0.95±0.03	36.65±3.62	0.96±0.03	34.23±3.12	0.96±0.04
	DuDoRNet	38.28±2.84	0.96±0.02	39.85±3.36	0.97±0.02	38.41±4.27	0.97±0.03
	Ours	39.41±2.46	0.96±0.01	41.31±3.22	0.98±0.01	39.82±4.46	0.98±0.02

values on two datasets with different under-sampling masks. As can be seen, the DuDoCAF yields the best results on all experiments. This indicates that our model is able to effectively fuse multi-contrast images which boosts the reconstruction performance. It is worth noting that YNet and UF-T2 reconstructions are far less than MINet results, which indicates that it is optimal to learn the interaction between two different contrast images step by step. More importantly, the DuDoRNet shows the second-best results demonstrates the powerful reconstruction ability of dual-domain learning.

Ablation Study Two key components are estimated, including: dual-domain (DD) learning and Cross-Attention Fusion (CAF) with reference image prior (RP) passing to the network. As shown in Table 2, (A) Baseline represents the network only consists of Residual Reconstruction Transformer block in the k-space domain. (B) *w/o* CAF adds DD learning to the Baseline network. (C)

Table 2: Ablative study on different settings of DuDoCAF under 8× acceleration on knee dataset. The best and second-best result are marked in red and blue, respectively.

Methods	CAF	DD	RP	Random(8×)		Radial (10×)		Spiral (10×)	
				PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Baseline			✓	25.88±2.01	0.81±0.04	26.91±1.67	0.85±0.05	26.40±1.81	0.84±0.06
w/o CAF		✓	✓	26.42±1.94	0.83±0.05	27.18±1.89	0.85± 0.04	26.98±1.73	0.85±0.04
w/o DD	✓		✓	26.79±2.02	0.83±0.04	27.95±1.66	0.86± 0.04	27.45±1.78	0.84±0.04
Ours	✓	✓	✓	27.45±1.52	0.86±0.03	28.91±1.32	0.88±0.03	28.90±1.59	0.87±0.03

w/o DD adds the CAF block to the Baseline architecture. (D) Ours indicates combination of Baseline, DD and CAF module. This indicates that learning from both *k*-space and image domain is really important even when adopting the cross-attention fusion strategy. Besides, the reconstruction results of *w/o* CAF are worse than those *w/o* DD, which clarifies the importance of fusing two contrast image features and exploiting the deep correlation between them.

4 Conclusion

We proposed a novel dual-domain cross-attention fusion mechanism (DuDoCAF) with recurrent transformer for fast multi-contrast MR Imaging. Firstly, the CAF module is able to better fuse the features of the fully-sampled reference images and under-sampled target images. Besides, the residual-reconstruction transformer helps the network to extract more informative features of the image for the target-contrast image restoration. Furthermore, the adopt dual-domain recurrent learning strategy is helpful to obtain better reconstructed images and reduce artifacts. Extensive experimental results show that, under different sampling patterns and acceleration factors, our proposed method significantly outperforms other state-of-the-art methods. In the future, we will extend the DuDoCAF from single-coil to a multi-coil reconstruction.

Acknowledgements This work was supported in part by National Natural Science Foundation of China under Grant 61902338 and 62001120, the Shanghai Sailing Program (No. 20YF1402400), The Hong Kong Polytechnic University under Project of Strategic Importance (no. 1-ZE2Q).

References

1. Chen, C.F.R., Fan, Q., Panda, R.: Crossvit: Cross-attention multi-scale vision transformer for image classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 357–366 (2021) 2.2
2. Chen, W., Zhao, J., Wen, Y., Xie, B., Zhou, X., Guo, L., Yang, L., Wang, J., Dai, Y., Zhou, D.: Accuracy of 3-t mri using susceptibility-weighted imaging to detect meniscal tears of the knee. Knee Surgery, Sports Traumatology, Arthroscopy **23**(1), 198–204 (2015) 1

3. Dar, S.U., Yurt, M., Shahdloo, M., Ildız, M.E., Tınaz, B., Çukur, T.: Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks. IEEE Journal of Selected Topics in Signal Processing **14**(6), 1072–1087 (2020) 1
4. Do, W.J., Seo, S., Han, Y., Ye, J.C., Choi, S.H., Park, S.H.: Reconstruction of multicontrast mr images through deep learning. Medical physics **47**(3), 983–997 (2020) 1, 2e, 3
5. Feng, C.M., Fu, H., Yuan, S., Xu, Y.: Multi-contrast mri super-resolution via a multi-stage integration network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 140–149. Springer (2021) 1, 2i, 3
6. Feng, C.M., Yan, Y., Chen, G., Fu, H., Xu, Y., Shao, L.: Accelerated multi-modal mr imaging with transformers. arXiv preprint arXiv:2106.14248 (2021) 3
7. Feng, C.M., Yan, Y., Fu, H., Chen, L., Xu, Y.: Task transformer network for joint mri reconstruction and super-resolution. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 307–317. Springer (2021) 1
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (Poster) (2015) 3
9. Knoll, F., Zbontar, J., Sriram, A., Muckley, M.J., Bruno, M., Defazio, A., Parente, M., Geras, K.J., Katsnelson, J., Chandarana, H., et al.: fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. Radiology: Artificial Intelligence **2**(1), e190007 (2020) 3
10. Korkmaz, Y., Dar, S.U., Yurt, M., Özbey, M., Cukur, T.: Unsupervised mri reconstruction via zero-shot learned adversarial transformers. IEEE Transactions on Medical Imaging (2022) 1
11. Liu, X., Wang, J., Sun, H., Chandra, S.S., Crozier, S., Liu, F.: On the regularization of feature fusion and mapping for fast mr multi-contrast imaging via iterative networks. Magnetic resonance imaging **77**, 159–168 (2021) 1
12. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014) 1
13. Sachan, T., Pinnaparaju, N., Gupta, M., Varma, V.: Scate: shared cross attention transformer encoders for multimodal fake news detection. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 399–406 (2021) 2.2
14. Sun, H., Cleary, J.O., Glarin, R., Kolbe, S.C., Ordidge, R.J., Moffat, B.A., Pike, G.B.: Extracting more for less: multi-echo mp2rage for simultaneous t1-weighted imaging, t1 mapping, mapping, swi, and qsm from a single acquisition. Magnetic resonance in medicine **83**(4), 1178–1191 (2020) 1
15. Sun, L., Fan, Z., Fu, X., Huang, Y., Ding, X., Paisley, J.: A deep information sharing network for multi-contrast compressed sensing mri reconstruction. IEEE Transactions on Image Processing **28**(12), 6141–6153 (2019) 1
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017) 2.2
17. Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., Shen, D.: Deep-learning-based multi-modal fusion for fast mr reconstruction. IEEE Transactions on Biomedical Engineering **66**(7), 2105–2114 (2018) 1, 2g, 3

18. Xu, Y., Zhao, H., Zhang, Z.: Topicaware multi-turn dialogue modeling. In: The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21) (2021) 2.2
19. Xuan, K., Sun, S., Xue, Z., Wang, Q., Liao, S.: Learning mri k-space subsampling pattern using progressive weight pruning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 178–187. Springer (2020) 3
20. Yang, Y., Wang, N., Yang, H., Sun, J., Xu, Z.: 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 Computer-Assisted Intervention. pp. 188–198. Springer (2020) 1
21. Zhou, B., Zhou, S.K.: 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. pp. 4273–4282 (2020) 1, 2k, 3