Dynamic Hybrid Unrolled Multi-scale Network for Accelerated MRI Reconstruction

Xiao-Xin Li[1,2], Fang-Zheng Zhu[1], Junwei Yang[5], Yong Chen[6], and Dinggang Shen[2,3,4](✉)

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

[2] School of Biomedical Engineering and State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
dinggang.shen@gmail.com

[3] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200232, China

[4] Shanghai Clinical Research and Trial Center, Shanghai 201210, China

[5] Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

[6] Department of Radiology, Case Western Reserve University, Cleveland, OH, USA

Abstract. In accelerated magnetic resonance imaging (MRI) reconstruction, the anatomy of a patient is recovered from a set of undersampled measurements. Currently, unrolled hybrid architectures, incorporating both the beneficial bias of convolutions with the power of Transformers have been proven to be successful in solving this ill-posed inverse problem. The multi-scale strategy of the intra-cascades and that of the inter-cascades are used to decrease the high compute cost of Transformers and to rectify the spectral bias of Transformers, respectively. In this work, we proposed a **d**ynamic **H**ybrid **U**nrolled **Mu**lti-**S**cale Network (dHUMUS-Net) by incorporating the two multi-scale strategies. A novel Optimal Scale Estimation Network is presented to dynamically create or choose the multi-scale Transformer-based modules in all cascades of dHUMUS-Net. Our dHUMUS-Net achieves significant improvements over the state-of-the-art methods on the publicly available fastMRI dataset.

Keywords: MRI acceleration · Unrolled architecture · Multi-scale strategy · Dynamic network

1 Introduction

Magnetic Resonance Imaging (MRI) as a non-invasive approach is a powerful diagnostic tool compared to competing modalities like CT or X-Rays. However, the MRI data acquisition process is inherently slow. Accelerated MRI [17]

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-72104-5_26.

M. G. Linguraru et al. (Eds.): MICCAI 2024, LNCS 15007, pp. 264–274, 2024.
https://doi.org/10.1007/978-3-031-72104-5_26

addresses this challenge by taking fewer measurements in k-space and thus reducing the time patients need to spend in the scanner. However, recovering the underlying anatomy from undersampled data is an ill-posed problem as measurements are less than unknowns.

Recently, Deep Neural Networks (DNNs) have been successfully used in image segmentation [3,20], registration [5], and reconstruction [2,11]. Compared to traditional compressed sensing methods, DNNs enabled higher quality reconstruction under higher acceleration rates. Most of the advanced DNN methods [7,13,21,22,25,28,30,32] adopt the deep unrolled architecture (DUA). A typical DUA is composed of several sequentially cascaded subnets, also called *cascades* [22], each of which ends with a data consistency (DC) layer [21] and acts as a refinement step to the final reconstruction [14]. The DC layers are very important to avoid losing or corrupting the sampled k-space data in the inputs after long-distance forward mapping. Therefore, the cascades in DUAs are usually very small, focusing only on the local features and having limited receptive fields [9].

With the emergence of Transformers [15,16], the limited receptive-field issue of unrolled models are well addressed [4,6,9,10,33]. However, MRI images have typically significantly higher dimension (e.g., 640×368 in [31]) than commonly used images, posing a significant challenge to contemporary Transformer-based models. As far as we know, HUMUS-Net [4] is the only Transformer-based model fully considering the issue incurred by the large-size input. HUMUS-Net extracts local-range features via using convolutions in high-dimensional space whereas extracts the long-range features via the Transformer blocks in low-dimensional space. To make the Transformer blocks perform well in low-dimensional space as far as possible, a U-Net-like multi-scale encoder-decoder architecture is adopted. Under this architecture, the large-size input can be progressively downsampled in the encoder path, and the multi-scale outputs in the encoder path can be passed to the decoder via long skip connections. This design can avoid information loss caused by downsampling as much as possible.

Note that ReconFormer [6], a very recent Transformer-based unrolled method, also utilizes a multi-scale strategy for MRI reconstruction. However, ReconFormer downsamples the input image by using only one step large-stride convolution in a cascade. Thus, the downsampling rate is limited. Moreover, ReconFormer varies the feature scales between cascades in a coarse-to-fine pyramid manner. The pyramid structure was first applied in PC-RNN [1], the champion model of the 2019 fastMRI Challenge, and is very import to rectify the low-frequency bias [19] of convolutions.

While the multi-scale strategy used in HUMUS-Net [4] lies in the *intra*-cascades in a U-shape style, the multi-scale strategy used in ReconFormer [6] or PC-RNN [1] lies in the *inter*-cascades in a pyramid manner. The two multi-scale strategies have their own benefits but have not been combined together for Transformer-based MRI reconstruction. In this work, we redesign the HUMUS-Net by incorporating the two multi-scale strategies and propose a **d**ynamic **H**ybrid **U**nrolled **Mu**lti-**S**cale Network (dHUMUS-Net) for accelerated MRI reconstruction[1]. The main challenge of using the pyramid structure is how to estimate the optimal scale for each of the cascades. The existing works [1,6] manually set the scales for all cascades according to experiences. This might lead to sub-optimal performance when the acceleration rate or the dataset changes. We will show that the optimal scale of a cascade mainly depends on the level of the repeated features, or the *repetition level* (RL), in the input image, and a large RL tends to require a large downsampling scale. We present a novel Optimal Scale Estimation Network (OSEN) to estimate the optimal scale of a given cascade as per the RL of the input image. Our dHUMUS-Net is designed by putting OSEN ahead each cascade of HUMUS-Net. Thus, the utilized Transformer modules can be constructed dynamically according to the prediction of the OSEN. To the best of our knowledge, this is the first attempt to design a dynamic unrolled structure for MRI reconstruction. We show through experiments on the fastMRI dataset that dHUMUS-Net yields higher fidelity reconstructions.

Our contributions are as follows: *1)* we propose a dHUMUS-Net by integrating the merits of the multi-scale strategy of the intra-cascades [4] and that of the inter-cascades [1,6]; *2)* we explore the dynamic construction of the unrolled Transformer-convolutional hybrid architecture for better reconstruction quality; and *3)* we perform extensive experiments using our model on the fastMRI dataset and obtain new state-of-the-art results for both the knee and the brain MRIs.

2 Problem Formulation

An MRI scanner obtains measurements of the patient anatomy in k-space via various receiver coils. The fully sampled k-space data can be obtained via $\mathrm{y} = \mathcal{A}(\mathrm{x})$, where $\mathrm{x} \in \mathbb{C}^n$ is the underlying patient anatomy of interest and usually has a very high dimension, and $\mathcal{A}$ is the linear forward operator [31] that first multiplies by the sensitivity maps and then applies 2D Fourier Transform (FT). Note that for simplicity, the measurement noise in the forward mapping is omitted. The anatomy image x can be recovered by $\mathrm{x} = \mathcal{A}^*(\mathrm{y})$, where $\mathcal{A}^*$ first applies 2D Inverse Fourier Transform (IFT) and then uses the reduce operator [22] to combine all individual coil images.

To accelerate MRI, only partial k-space data y_u is acquired $\mathrm{y}_u = M_u\mathcal{A}(\mathrm{x})$,where M_u is a diagonal matrix representing a binary undersampling mask for $u\times$ acceleration. As y_u is highly undersampled, directly applying $\mathcal{A}^*$ on y_u will lead to a highly aliased reconstruction $\mathrm{x}_u = \mathcal{A}^*(\mathrm{y}_u)$. Deep unrolled architectures (DUAs) [2,4,6,10,14] are widely used to further perform reconstruction from x_u. However, x_u is usually high dimensional and can lead to high compute cost.

Fortunately, we observe that x_u is *highly compressible*. This is because x_u is obtained from the zero-filled undersampled k-space data via using IFT. After IFT, each zero value in k-space can spread over all pixels of x_u and thus results

[1] Here, "Hybrid" means the a Transformer-convolutional hybrid operations. Since the work of Xiao *et al.* [27], Transformers have been bound with convolutions for vision tasks.

in a lot of repeated features in image domain. Such a compressibility can extend to the inputs of the intermediate cascades. In this work, we will explore how to use the compressibility of the aliased input images and the multi-scale strategies of the intra- and inter-cascades to boost the reconstruction performance and efficiency of the Transformer-based DUAs.

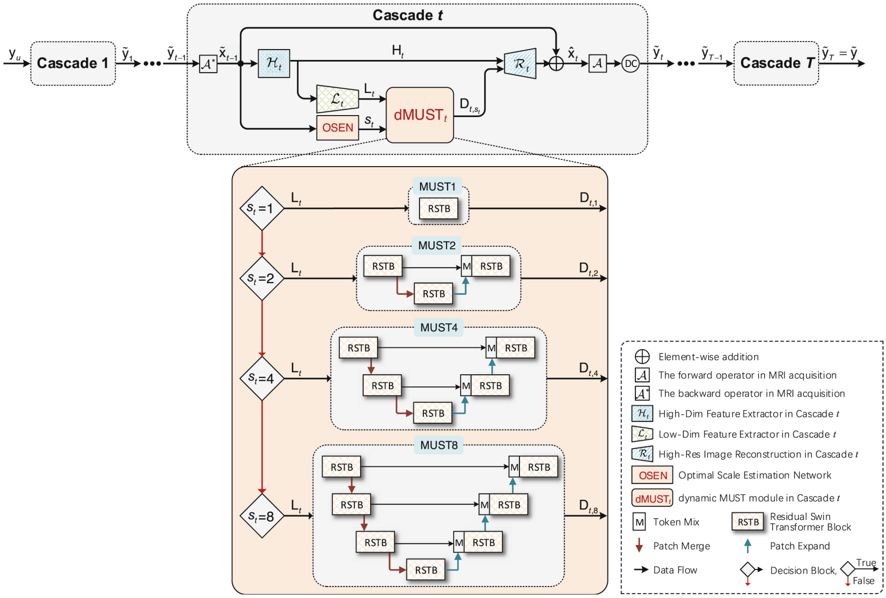
Cascade t
y_u
Cascade 1
ȳ_1
ȳ_{t−1}
A*
x̄_{t−1}
H_t
H_t
L_t
L_t
OSEN
s_t
dMUST_t
D_{t,s_t}
R_t
x̂_t
A
DC
ȳ_t
ȳ_{T−1}
Cascade T
ȳ_T = ȳ
s_t=1
MUST1
RSTB
D_{t,1}
s_t=2
MUST2
D_{t,2}
s_t=4
MUST4
D_{t,4}
s_t=8
MUST8
D_{t,8}
M
Element-wise addition
The forward operator in MRI acquisition
The backward operator in MRI acquisition
High-Dim Feature Extractor in Cascade t
Low-Dim Feature Extractor in Cascade t
High-Res Image Reconstruction in Cascade t
Optimal Scale Estimation Network
dynamic MUST module in Cascade t
Token Mix
Residual Swin Transformer Block
Patch Merge
Patch Expand
Data Flow
Decision Block,
True
False

Fig. 1. Overview of the proposed **d**ynamic **H**ybrid **U**nrolled **Mu**lti-**S**cale Network (dHUMUS-Net). The main module is the dMUST, i.e., the dynamic MUST (**Mu**lti-scale residual **S**win **T**ransformer), which consists of several MUST-based branches (e.g., MUST1, MUST2 and etc.), and will be dynamically constructed during training according to the repetition level of the input image estimated by OSEN.

3 Method

The network architecture of the proposed **d**ynamic **H**ybrid **U**nrolled **Mu**lti-**S**cale Network (dHUMUS-Net) is based on the HUMUS-Net [4] and illustrated in Fig. 1. Before delving into dHUMUS-Net, we first briefly introduce the main blocks of HUMUS-Net. HUMUS-Net consists of T cascades, namely HUMUS-Blocks, each of which consists of a high-dimension-feature extractor $\mathcal{H}$, a low-dimension-feature extractor $\mathcal{L}$, a deep-and-low-dimension-feature extractor, namely **Mu**lti-scale residual **S**win **T**ransformer (MUST), and a reconstruction operator $\mathcal{R}$. Note that $\mathcal{H}$, $\mathcal{L}$ and $\mathcal{R}$ are convolutional blocks (ConvBlocks), whereas MUST is a Transformer-convolutional hybrid module. The main process in Cascade t can be written as follows

$$\mathsf{H}_t = \mathcal{H}_t\left(\tilde{\mathrm{x}}_{t-1}\right) \tag{1}$$

$$\mathsf{L}_t = \mathcal{L}_t\left(\mathsf{H}_t\right) \tag{2}$$

$$\mathsf{D}_t = \mathrm{MUST}_t\left(\mathsf{L}_t\right) \tag{3}$$

$$\hat{\mathrm{x}}_t = \tilde{\mathrm{x}}_{t-1} + \mathcal{R}_t\left(\mathsf{H}_t, \mathsf{D}_t\right), \tag{4}$$

where $\tilde{\mathrm{x}}_{t-1} = \mathcal{A}^*\left(\tilde{\mathrm{y}}_{t-1}\right)$ is the reduced reconstruction result of Cascade $t-1$ and we use the subscript t to denote the outputs or operators from Cascade t.

As shown in Fig. 1, dHUMUS-Net differs from HUMUS-Net by replacing the operation of Eq. (3) with

$$s_t = \mathrm{OSEN}\left(\tilde{\mathrm{x}}_{t-1}\right) \tag{5}$$

$$\mathsf{D}_{t,s_t} = \mathrm{dMUST}_t\left(s_t, \mathsf{L}_t\right). \tag{6}$$

We next illustrate dMUST and OSEN, respectively.

dMUST (Dynamic MUST). HUMUS-Net fixes the configuration of the MUST module but neglects that the *repetition level* (RL) of the inputs to the cascades might reduce with the increase of the cascading depth. This requires the maximum downsampling scale, called *max-scale*, of the MUST module should be also adjusted to adapt to the input data. Accordingly, we propose the dMUST module, which can be defined as follows

$$\mathrm{dMUST}\left(s, \mathsf{L}\right) \triangleq \begin{cases} \mathrm{MUST1}\left(\mathsf{L}\right) & \text{if } s = 1 \\ \mathrm{MUST2}\left(\mathsf{L}\right) & \text{if } s = 2 \\ \mathrm{MUST4}\left(\mathsf{L}\right) & \text{if } s = 4 \\ \mathrm{MUST8}\left(\mathsf{L}\right) & \text{if } s = 8 \end{cases}, \tag{7}$$

where MUSTs denotes the MUST module with the max-scale s. Note that the max-scale of a MUST module can only be 2^n (n is a non-negative integer) due to the definition of the downsampling operator, PatchMerge [4], used in the encoder path. In Eq. (7), we limit the candidate max-scales to be in the set $\mathcal{S}_{\mathrm{MS}} = \{1, 2, 4, 8\}$. The max-scales in $\mathcal{S}_{\mathrm{MS}}$ is set to be ≤ 8 because for the highest 8× acceleration rate used in this work, setting max-scales ≤ 8 is enough to reduce the RL of the input images.

OSEN (Optimal Scale Estimation Network). Given an input image, OSEN uses its RL to estimate the optimal downsampling scale that can be used by dMUST to dynamically create or choose a MUST branch. The mapping relationship between the RL and the optimal scale is non-trivial. Inspired by the universal approximation power of DNNs [8], we propose to design the mapping function as a neural network. Suppose we have a set of training samples $\mathit{\Gamma} = \left\{ \left(\mathrm{x}_u^{(i)}, \mathrm{x}^{(i)} \right)\Big|_{u=r/2}^{r} \right\}_{i=1}^{N}$, where r is the target acceleration rate, and the acceleration rates of the training samples range from $r/2$ to r to ensure that the possible RLs of the outputs of all cascades can be covered as far as possible. For training, we should further quantify the RL and label the optimal scale for each of the training sample in Γ.

Quantify the RL. As the repeated features are globally distributed in the whole image, we decompose the input image into several sub-images by using the pixel-unshuffle (PU) operator [26] and measure the RL by using the SSIM-based similarities between all PU outputs. Given a PU-factor s ($s \geq 2$), the PU operator can produce s^2 outputs and thus leads to a similarity vector v_s consisting of $\frac{1}{2}s^2\left(s^2-1\right)$ values. Given a PU-factor set $\mathcal{S}_{\mathrm{PU}}$, we represent the RL of an input image as follows

$$v_{\mathrm{RL}} = \left[v_2^\top, v_3^\top, \cdots, v_{|\mathcal{S}_{\mathrm{PU}}|}^\top\right]^\top. \tag{8}$$

Label the Optimal Scale. We first train four HUMUS-Blocks [4], whose MUST modules are set to MUST1, MUST2, MUST4, and MUST8, respectively, by using the training set Γ. We then label each of the training samples as follows

$$s_{\mathrm{Opt}}^{(i)} = \arg\max_{s \in \mathcal{S}_{\mathrm{MS}}} \mathrm{SSIM}\left(\mathrm{HUMUS}s\left(\mathrm{x}_u^{(i)}\right), \mathrm{x}^{(i)}\right), \tag{9}$$

where HUMUSs denotes the HUMUS-Block configured with MUSTs, and SSIM (Structural Similarity Index Measure) is used to evaluate the reconstruction performance. Note that the MUST modules used in HUMUSs are independent of those used in dHUMUS-Net, where all MUST modules used in different cascades should be retrained in an end-to-end manner.

We now have the final training set $\left\{\left(v_{\mathrm{RL}}^{(i)}, s_{\mathrm{Opt}}^{(i)}\right)\right\}_{i=1}^{N\cdot(r/2+1)}$. As the similarity vectors v_s in $v_{\mathrm{RL}}^{(i)}$ can be considered as a 1D time series, we use an RNN (Recurrent Neural Network) to model the mapping from v_{RL} to s_{Opt}.

The design of dMUST optimizes the scale configuration for different inputs, making different inputs having different optimal scales go along different branches of dMUST. However, different branches of dMUST share the same ConvBlocks in each cascade. This design can not only save network parameters but also boost the filter diversity due to the benefit of multi-scale training [12]. As well known, filter diversity [23,24] is very important to enhance performance.

4 Experiments

4.1 Datasets and Experimental Setting

We use the knee and brain datasets from fastMRI competition [31]. The knee dataset has two MRI modalities: proton-density weighting with (PDFS) and without (PD) fat suppression, and includes single-coil and multi-coil tasks with 973 volumes (34,742 slices) for training and 199 volumes (7,135) for validation. The raw k-space data are with the matrix of size 640 × 368 or 640 × 372. The brain dataset has four MRI modalities, namely T1, T1POST, T2, and FLAIR, and only includes multi-coil task with 4,469 volumes (70,748 slices) for training and 1,378 volumes (21,842 slices) for validation. The raw k-space data have different sizes, e.g., 768 × 396 and 640 × 272. In this work, we only the multi-coil datasets. We use the Sensitivity Map Estimation (SME) module of E2E-VN [22] to estimate the coil sensitivity maps used in the forward and reduce operators in our dHUMUS-Net.

Acceleration rates (AR) are set to 4 and 8. We use equispaced undersampling masks as they are easier to implement in MRI scanners [22]. The undersampled k-space data and the zero-filled images are produced according to the illustration in Sect. 2. The experimental results on different modalities are reported separately so that we can see the effects of our method on different tasks clearly. We use the SSIM as the evaluation metric. Other implementation details can be found in the *supplementary*.

4.2 Ablation Study

Two max-scale configuration strategies for dHUMUS-Net should be compared: fixed configuration (FC) and OSEN-based dynamic configuration (DC). For FC, we choose the four max-scales from $\mathcal{S}_{\text{MS}}$. For DC, we need to consider the mini-batch generation strategy, which suggests that the samples in the two parts, $\mathcal{B}_s^{(1)}$ and $\mathcal{B}_s^{(2)}$, of a mini-batch $\mathcal{B}_s$ should satisfy a certain ratio. We use OSEN($\mathcal{B}_s^{(1)}$:$\mathcal{B}_s^{(2)}$), specifically, OSEN(1:0) and OSEN(7:3), to represent how the mini-batch generation strategy cooperate with OSEN.

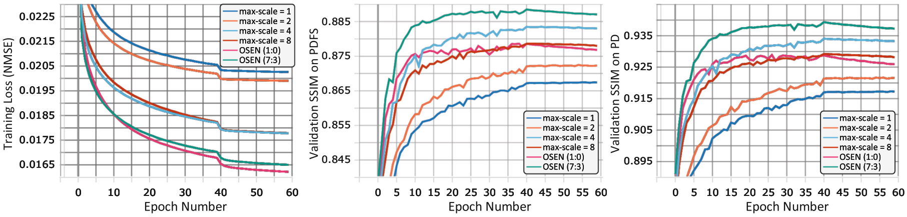

Fig. 2. Ablation study for dHUMUS-Net under various max-scale configurations. We use the fastMRI multi-coil knee dataset and AR = 8.

Figure 2 plots the comparison results. For FC, we can see from Fig. 2(b) and (c) that max-scale = 4, instead of max-scale = 8, leads to the best performance. It means that manually setting the max-scales has the risk to lead to sub-optimal performance. Using OSEN(7:3) rather than OSEN(1:0) can achieve the best performance on both PDFS and PD. We can see from Fig. 2(a) that OSEN(1:0) leads to the best training loss but bad generalization performance. This means that separating the data samples with different optimal scales into different mini-batches might make the network performance bias to the data whose optimal scales are predominant.

4.3 Comparison Study

Table 1. Evaluation results on the validation dataset of fastMRI multi-coil tasks.

AR	Method	Knee		Brain			
		PDFS	PD	T1	T1POST	T2	FLAIR
4	PC-RNN	0.9014 ± 0.010	0.9472 ± 0.015	0.9632 ± 0.018	0.9651 ± 0.019	0.9603 ± 0.016	0.9205 ± 0.050
	ReconFormer	0.9208 ± 0.013	0.9635 ± 0.016	0.9673 ± 0.018	0.9694 ± 0.017	0.9652 ± 0.023	0.9302 ± 0.046
	HUMUS-Net	0.9170 ± 0.015	0.9578 ± 0.012	0.9642 ± 0.019	0.9650 ± 0.019	0.9635 ± 0.015	0.9273 ± 0.047
	dHUMUS-Net	**0.9286 ± 0.014**	**0.9693 ± 0.011**	**0.9723 ± 0.016**	**0.9763 ± 0.018**	**0.9695 ± 0.019**	**0.9373 ± 0.035**
8	PC-RNN	0.8623 ± 0.027	0.9125 ± 0.021	0.9401 ± 0.015	0.9532 ± 0.022	0.9422 ± 0.030	0.8984 ± 0.059
	ReconFormer	0.8745 ± 0.023	0.9218 ± 0.023	0.9433 ± 0.020	0.9546 ± 0.023	0.9433 ± 0.030	0.9056 ± 0.060
	HUMUS-Net	0.8827 ± 0.023	0.9221 ± 0.022	0.9490 ± 0.019	0.9636 ± 0.018	0.9510 ± 0.027	0.9182 ± 0.055
	dHUMUS-Net	**0.8894 ± 0.022**	**0.9355 ± 0.019**	**0.9530 ± 0.013**	**0.9726 ± 0.015**	**0.9592 ± 0.019**	**0.9248 ± 0.048**

We compare the proposed dHUMUS-Net with three state-of-the-art methods based on the unrolled architecture: a CNN-based method, namely PC-RNN [1], and two Transformer-based methods, including HUMUS-Net [4] and ReconFormer [6]. The source codes of these methods, except for PC-RNN, can be available from the websites of the authors. We implemented PC-RNN based on the source code of CRNN [18].

Table 1 compares the reconstruction performance of the four compared methods for 4× and 8× accelerations. It is interesting to note that HUMUS-Net is outperformed by ReconFormer when AR $=4$ but performs better than ReconFormer when AR $=8$. This is mainly because the repetition levels (RLs) of the input data under AR $=4$ are much lower than those under AR$=8$ and the large max-scales used in HUMUS-Net tend to cause information loss for inputs with low RLs. Our dHUMUS-Net achieves the best performance on all modalities of the two fastMRI multi-coil tasks. This can be attributed to the inclusion of the dynamic modules, OSEN and dMUST, to dynamically adapt to different input MRI images with different RLs. Also due to the dynamic design, our method can run well on a machine with 11 GB GPU memory, while HUMUS-Net requires $\geq$16 GB GPU memory.

Figure 3 visually compared the reconstruction quality of the compared methods for 8× accelerations. As can be seen, our dHUMUS-Net can restore more fine-scaled information and anatomy structures than the other methods. Also, dHUMUS-Net can well remove the aliasing artifacts produced by HUMUS-Net (see the image patches in the large dotted rectangle in Fig. 3A).

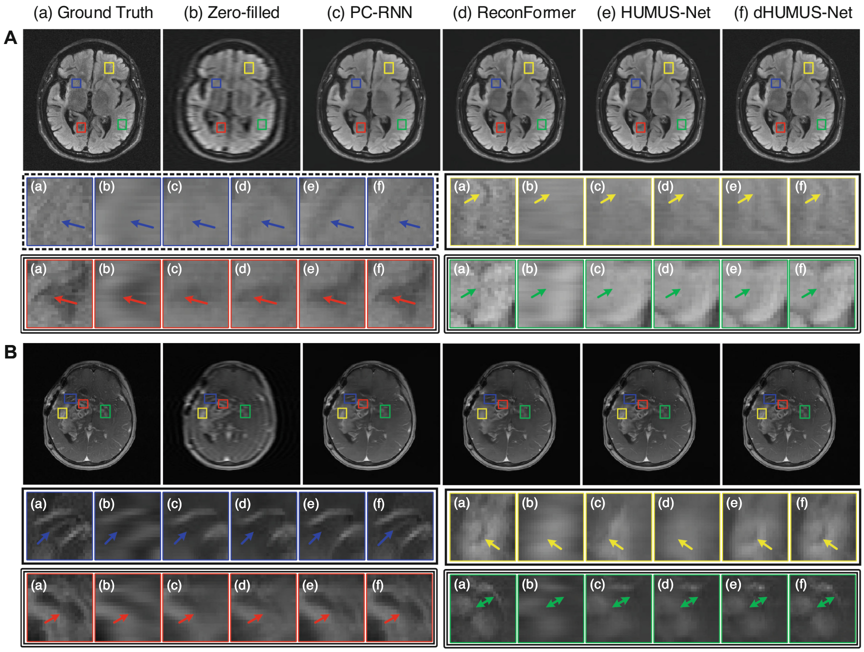

Fig. 3. Visual comparison of the reconstructed images for AR = 8. A/B: FLAIR/T1POST of brain. The image patches grouped in the large rectangles with doted lines, solid lines, and double lines, respectively, show that how the aliasing artifacts, the very fine-scaled information, and the anatomical structures are dealt with by the compared methods. More visual comparison can be found in the *supplementary*.

5 Conclusions

We provide a dHUMUS-Net to resolve the high dimension and high repetition level (RL) issues in MRI reconstruction by incorporating the intra- and inter-cascade multi-scale strategies into the unrolled Transformer-convolutional hybrid architecture. OSEN and dMUST are designed for dynamic module selection according to the RL of the input data. Experiments on the fastMRI dataset demonstrated the effectiveness of our method. Future work will focus on better design of the dMUST and make dMUST can be configured with more max-scales beyond 2^n so that the model can better adapt to data. We will also study the adaption of OSEN on other well-known convolutions-based [1,22,28,29] and Transformers-based [6,10] unrolled architectures.

Acknowledgments. This study was funded by Zhejiang Provincial Natural Science Foundation of China (Grant No. LGF22F020027) and by National Natural Science Foundation of China (Grant No. 62271448 and 62373324).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, E.Z., Wang, P., Chen, X., Chen, T., Sun, S.: Pyramid convolutional RNN for MRI image reconstruction. IEEE Trans. Med. Imaging **41**(8), 2033–2047 (2022)
2. Chen, Y., et al.: AI-based reconstruction for fast MRI-a systematic review and meta-analysis. Proc. IEEE **110**(2), 224–245 (2022)
3. Dolz, J., Desrosiers, C., Wang, L., Yuan, J., Shen, D., Ayed, I.B.: Deep CNN ensembles and suggestive annotations for infant brain MRI segmentation. Comput. Med. Imaging Graph. **79**, 101660 (2020)
4. Fabian, Z., Tinaz, B., Soltanolkotabi, M.: HUMUS-Net: hybrid unrolled multi-scale network architecture for accelerated MRI reconstruction. Adv. Neural. Inf. Process. Syst. **35**, 25306–25319 (2022)
5. Fan, J., Cao, X., Xue, Z., Yap, P.-T., Shen, D.: Adversarial similarity network for evaluating image alignment in deep learning based registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 739–746. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_83
6. Guo, P., Mei, Y., Zhou, J., Jiang, S., Patel, V.M.: ReconFormer: accelerated MRI econstruction using recurrent Transformer. IEEE Trans. Med. Imaging **43**(1), 582–593 (2024)
7. Hammernik, K., et al.: Learning a variational network for reconstruction of accelerated MRI data. Magn. Reson. Med. **79**(6), 3055–3071 (2018)
8. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. **2**(5), 359–366 (1989)
9. Hu, D., Zhang, Y., Zhu, J., Liu, Q., Chen, Y.: TRANS-Net: transformer-enhanced residual-error alternative suppression network for MRI reconstruction. IEEE Trans. Instrum. Meas. **71**, 1–13 (2022)
10. Huang, J., et al.: Swin transformer for fast MRI. Neurocomputing **493**, 281–304 (2022)
11. Hyun, C.M., Kim, H.P., Lee, S.M., Lee, S., Seo, J.K.: Deep learning for undersampled MRI reconstruction. Phys. Med. Biol. **63**(13), 135007 (2018)
12. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
13. Li, X.-X., Chen, Z., Lou, X.-J., Yang, J., Chen, Y., Shen, D.: Multimodal MRI acceleration via deep cascading networks with peer-layer-wise dense connections. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 329–339. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_32
14. Liang, D., Cheng, J., Ke, Z., Ying, L.: Deep magnetic resonance image reconstruction: inverse problems meet neural networks. IEEE Signal Process. Mag. **37**(1), 141–151 (2020)
15. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: SwinIR: image restoration using Swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833–1844 (2021)
16. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)
17. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. **58**(6), 1182–1195 (2007)

18. Qin, C., Schlemper, J., Caballero, J., Price, A.N., Hajnal, J.V., Rueckert, D.: Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging **38**(1), 280–290 (2019)
19. Rahaman, N., et al.: On the spectral bias of neural networks. In: Proceedings of the International Conference on Machine Learning, pp. 5301–5310 (2019)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
21. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic MR image reconstruction. IEEE Trans. Med. Imaging **37**(2), 491–503 (2018)
22. Sriram, A., et al.: End-to-end variational networks for accelerated MRI reconstruction. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 64–73. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_7
23. Wang, J., Chen, Y., Chakraborty, R., Yu, S.X.: Orthogonal convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11502–11512 (2020)
24. Wang, Q., Guo, G.: DSA-face: diverse and sparse attentions for face recognition robust to pose variation and occlusion. IEEE Trans. Inf. Forensics Secur. **16**, 4534–4543 (2021)
25. Wang, S., et al.: DeepcomplexMRI: exploiting deep residual network for fast parallel MR imaging with complex convolution. Magn. Reson. Imaging **68**, 136–147 (2020)
26. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-ESRGAN: training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1905–1914 (2021)
27. Xiao, T., Singh, M., Mintun, E., Darrell, T., Dollár, P., Girshick, R.: Early convolutions help transformers see better. Adv. Neural. Inf. Process. Syst. **34**, 30392–30400 (2021)
28. Yang, Y., Sun, J., Li, H., Xu, Z.: ADMM-CSNet: a deep learning approach for image compressive sensing. IEEE Trans. Pattern Anal. Mach. Intell. **42**(3), 521–538 (2020)
29. Yiasemis, G., Sonke, J.J., Sśnchez, C., Teuwen, J.: Recurrent variational network: a deep learning inverse problem solver applied to the task of accelerated MRI reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 722–731 (2022)
30. You, D., Xie, J., Zhang, J.: ISTA-Net++: flexible deep unfolding network for compressive sensing. In: Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 1–6 (2021)
31. Zbontar, J., et al.: fastMRI: an open dataset and benchmarks for accelerated MRI. arXiv preprint arXiv:1811.08839 (2018)
32. Zhang, J., Ghanem, B.: ISTA-net: interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1828–1837 (2018)
33. Zhou, B., et al.: DSFormer: a dual-domain self-supervised transformer for accelerated multi-contrast MRI reconstruction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 4966–4975 (2023)

Supplementary Materials

Xiao-Xin Li[1,2], Fang-Zheng Zhu[1], Junwei Yang[5], Yong Chen[6], and Dinggang Shen[2,3,4,*]

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
[2] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[3] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200232, China
[4] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
[5] Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
[6] Department of Radiology, Case Western Reserve University, Cleveland, Ohio, USA

1 Implementation Details

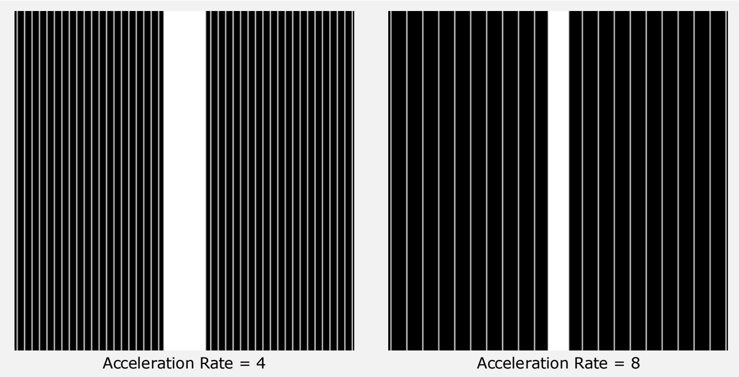

Fig. 1. The equispaced undersampling masks we used.

We adopt the equispaced undersampling masks [1] for 4× and 8× accelerations, as shown in Fig. 1. The equispaced undersampling masks continuously sample partial low-frequency lines from the center of k-space and uniformly sample high-frequency lines from the remaining k-space.

We use the lookahead version of Adam optimizer [2] with a momentum of 0.9 and a weight decay of 10^{-4} and perform 60 epochs in the training stage. The learning rate is initially set to 10^{-4} and dropped by a factor of 10 at epoch 40. Mini-batch-based training was adopted. We set the number of cascades $T = 8$, the size of local windows to be 8, and MLP ratio to be 2. All RSTBs consist of 2 STLs with embedding dimension of 66. We implement the proposed model using PyTorch on NVIDIA RTX5000 24GB GPUs.

Performance comparison tables report metrics' mean and standard deviation across test subjects. We evaluate the statistical significance of performance differences using Wilcoxon sign-rank tests.

2 Supplemental Experiments

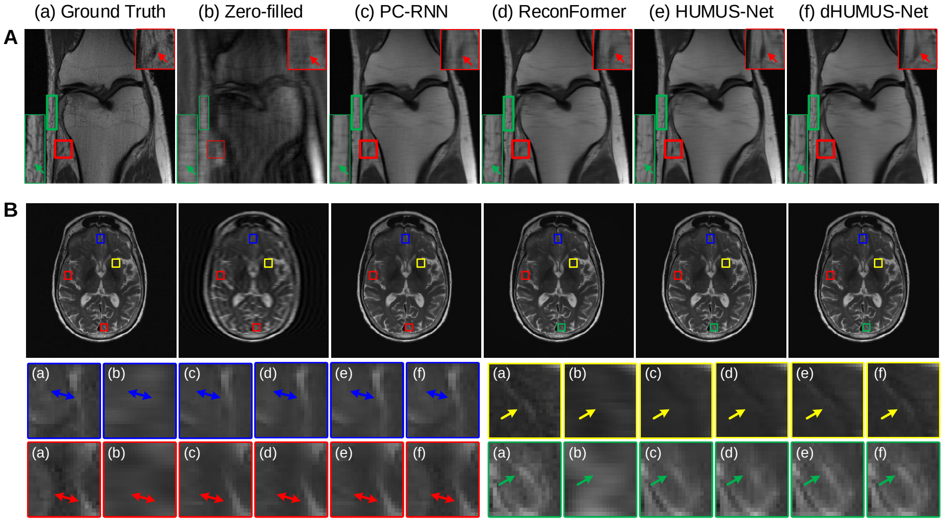

Fig. 2. Visual comparison of the reconstructed images for AR=8. A: PD of knee, B: T2 of brain.

References

1. Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C.L., Yakubova, N., Knoll, F., Johnson, P.: End-to-end variational networks for accelerated MRI reconstruction. In: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention. pp. 64–73. Springer International Publishing, Cham (2020)
2. Zhang, M., Lucas, J., Ba, J., Hinton, G.E.: Lookahead optimizer: k steps forward, 1 step back. Advances in neural information processing systems **32** (2019)