High-Fidelity MRI Reconstruction Using Adaptive Spatial Attention Selection and Deep Data Consistency Prior

High-Fidelity MRI Reconstruction Using Adaptive Spatial Attention Selection and Deep Data Consistency Prior

Jingshuai Liu, Chen Qin, and Mehrdad Yaghoobi

***Abstract*—Compressed sensing (CS) has shown great potential for fast magnetic resonance imaging (fastMRI). Traditional CS methods model the inverse problem by leveraging the sparsity prior to guarantee the success of signal recovery, which is not rich enough to capture the detailed features of MRI modality. The other challenge is computational complexity in CS methods which often include an iterative optimization-based solver, hindering the growth and development of modern high resolution MRI. Inspired by existing researches in machine vision tasks, two novel network blocks are presented here which respectively leverage a) the spatial correlations and b) data consistency prior, and a novel multi-level densely connected framework is devised to improve the model capacity for removing aliasing artifacts from the under-sampled MR images and recovering missing anatomical information in high resolution MRIs. It is demonstrated that the framework produces more realistic and faithful structures and textural details, providing superior reconstructions in terms of less visual artifacts and relevant metrics.**

***Index Terms*—GAN-based framework, MRI reconstruction, adaptive spatial attention**

I. Introduction

SINCE its advent in the last century, magnetic resonance imaging (MRI) provides a radiation-free and non-invasive imaging tool and has revolutionized medical imaging and radiology. However, known as a slow imaging modality, MRI suffers from enormous consumption in acquisition and reconstruction time, which hinders its application in time-critic diagnosis. Various methods have been proposed to facilitate the time-consuming acquisition and reconstruction steps, such as parallel imaging (PI) which makes use of the sensitivity correlations to recover clean images from the under-sampled measurements [1]. Those methods can be burdened with expensive and complex equipments and high computational complexity, and are difficult to remove strong aliasing artifacts, using traditional reconstruction methods.

Model-based methods solve the ill-posed inverse problems by leveraging the image prior in the form of data sparsity. Assuming that the targets can be expressed via sparse representations, e.g. in Fourier space, compressed sensing (CS) achieves accurate MRI reconstruction by solving nonlinear optimization with the sparsity prior as regularization. Different from [2] which assumes sparse representations in the image domain or [3]–[5] which require sparse signals in some transform domains, the work introduced in [6] shows that the predetermined sparsifying transform can be replaced by a learnable basis via dictionary learning to enable more parsimonious representations. However, the challenge in holding the sparsity hypothesis in real-world scenarios and the restricted capacity of the sparsity prior to capture complex structures put constraints on the development of CS methods in modern MRI, and may restrict the achievable resolution.

Many recent works leverage the representation power of neural networks to model the distribution of MR images and recover the corrupted signals. The problem of parallel MR imaging is tackled in [7] by using U-shaped networks (U-net) to predict the missing k-space data. A cascaded framework is introduced in [8] to exhibit residual learning of aliasing artifacts and attain de-aliased outputs. A variational auto-encoder (VAE) is used in [9] to perform the variational inference on reconstructions which yields promising results by maximizing the estimated posteriori (MAP). Generative adversarial networks (GAN) [10] have achieved great success in image generation and shown their potential to provide high-quality MRI reconstructions. A GAN-based framework is proposed in [11] to encourage sharp and realistic details in reconstructed images. The methods introduced in [12], [13] perform MRI reconstruction by optimizing in latent space of a pre-trained generative model. The model introduced in [14], dubbed GANCS, leverages the interleaved structure of the null space operation and multilayer residual blocks to explore the targeted manifold and explicitly ensure data consistency. Motivated by the previous works in image domain translation [15], adversarial cyclic frameworks have been proposed to recover the under-sampled MR images. Concurrent researches include Cycle-MedGAN [16] where the perceptual metric [17] is incorporated to encourage visual realism, and [18] which solves the inverse problems using cycle-consistent adversarial networks (CycleGAN).

To provide the guarantee of convergence for MRI reconstruction methods, model-based optimization can be incorporated with deep learning-based frameworks. A primal-dual framework is introduced in [19] which transforms the conventional L_1-based regularization in compressive sensing to the inner-product and provides superior results over other optimization-based network models, e.g. ADMM-Net [20] and IFR-Net [21], in CS-MRI problems. The work in [22]

J. Liu, C. Qin, and M. Yaghoobi are with the Institute for Digital Communications, School of Engineering, University of Edinburgh, EH9 3JE UK (e-mail: J.Liu@ed.ac.uk; Chen.Qin@ed.ac.uk; m.yaghoobi-vaighan@ed.ac.uk).

Manuscript received August 14, 2022.

exploits multi-scale transforms in a ISTA-net [23] framework, showing to significantly improve the model performance. An automatic feedback mechanism is introduced in [24] to guide the data-driven optimization cycle and provide more reliable treatments. It is shown in [25] that plug-and-play methods can potentially yield better reconstruction quality when appropriately combining neural networks with model-based optimizations.

In this paper, we introduce a novel framework to provide high-fidelity MRI reconstructions. An adaptive spatial attention selection module (ASASM) and deep data consistency block (DCB) based model is proposed to better recover missing information and preserve realistic structures and textures. To exploit useful information contained in the spatial regions of high-resolution features, the ASASM module is devised to efficiently capture contextual information and perform data-adaptive kernel prediction to increase the diversity of attention patterns without introducing huge computational complexity. Traditional data consistency methods inevitably introduce a bottleneck structure which potentially has a negative impact on the model performance. To address this issue, the DCB was devised which takes the advantage of residual learning in feature space to avoid the bottleneck design and improve the reconstruction performance. We show that the framework produces superior results in terms of image quality and reconstruction accuracy, compared with state-of-the-art methods. To verify the efficacy of our method, we conduct ablation studies which demonstrate the role of the suggested multi-level structure and network modules to register impressive performance gain for MRI reconstruction. The main contributions of this paper are summarized below:

- an optimization-inspired reconstruction framework which has the capacity to leverage domain-specific knowledge into the reconstruction pipeline to encourage more faithful results;
- a novel module to perform "spatial attention selection" which endows the network with spatial kernel diversities and adaptive properties;
- a deep "data consistency block" to enhance the reconstruction quality by tighter pairing predictions with the observed measurements;
- a "multi-level densely connected architecture" to enable feature transmission and reuse at different levels, and impressively boost performance;
- by simulations, it has been demonstrated that the proposed approach can achieve high-fidelity reconstructions under aggressive sampling rates, outperforming other deep-learning based methods in the relevant metrics.

II. Related Works

In this section, we give a brief review of current researches closely related to our method, and show how our approach is associated with their contributions.

A self-attention module is introduced in [26] to capture non-local dependencies, which is utilized in [27], [28] to improve the reconstruction quality. The work in [29] employs both frequency-attention and channel-attention blocks to enhance the model efficacy and better reconstruct MR images. A convolutional block attention module (CBAM) is proposed in [30] to perform adaptive feature refinement. It is suggested in [31] to integrate the spatial attention unit of CBAM to a GAN-based framework, which can more accurately recover anatomical structures for MR images. However, it is still a challenge to recover faithful structures and yield sophisticated textural details in aggressive under-sampling diets, which shows potential opportunities for performance enhancement using more effective attention designs. In a recent conference paper, we introduce a GAN-based framework which presents feature refinement and attentive selection to generate sharp and realistic textures in MRI reconstructions [32]. We however found that in some cases it can produce slight artifacts and fail to properly recover structural details. To address the drawbacks of the method in [32] and push the quality of reconstructions, we propose a novel attention module to perform spatial attention with diverse local priors for compressive MRI.

In recent works of MR image restoration, it is a growing trend to leverage model-based algorithms to provide convergence guarantee for end-to-end solutions and achieve high-quality MRI reconstructions. A deep framework is proposed in [33] to recover under-sampled inputs with data consistency modules to correct the reconstructed data. DeepCascade introduced in [34] reconstructs dynamic MRI sequences using a deep cascade of convolutional neural networks (CNN) and data consistency layers. Based on the iterative shrinkage-thresholding algorithm, ISTA-net is derived in [23] for image compressed sensing by replacing the pre-defined sparsifying transform with deep neural networks. However, the bottleneck design in basic data consistency blocks can potentially degrade the model performance. We use an efficient structure to explore the data consistency in feature space and benefit the model in inferring missing information from corrupted inputs.

As an efficient network structure, densely connected layers are proposed in [35] to prevent the gradient vanishing problem which hinders us from increasing the model depth to build high-performance networks. As displayed in Fig. 1 (a), each bypass connection concatenates the current feature maps to the consecutively reused feature volume. Based on the work in [35], a deep residual network is introduced in [36] to recover heavily down-sampled images. An enhanced model, dubbed ESRGAN, is derived in [37] to retrieve more natural textures, by leveraging the benefits of both residual and dense connections. Inspired by the success of feature reuse in vision tasks, we here incorporate the multi-level dense connections in a cascaded network architecture to enable easier information transmission and enhance the model capacity.

III. Method

We introduce our method for MRI reconstruction in this section. We first formalize the reconstruction problem and describe the design of the reconstruction pipeline. Subsequently, the devised network components used in our framework are introduced. In the last part of this section, we give the objective functions adopted in the training phase.

A. Problem Formulation

MRI reconstruction is traditionally posed as an optimization problem, with A denoting the encoding matrix, y the measurement, and $R(x)$ a regularization term,

$$\min_{m} \lambda \|Am - y\|^2 + R(m). \tag{1}$$

Compressed sensing (CS) methods adopt the sparsity prior as regularization and solve the problem using iterative algorithms, which are computationally expensive and challenging to recover highly under-sampled signals, due to the limited expressiveness of sparsity hypothesis. We propose to leverage the representation of deep neural networks and provide MRI reconstructions in an end-to-end diet. To solve (1), an auxiliary variable u is imported to obtain:

$$\min_{m,u} \lambda \|Am - y\|^2 + R(u),\ s.t.\, m = u, \tag{2}$$

The unconstrained form is derived with the Lagrangian α:

$$\min_{m,u} \lambda \|Am - y\|^2 + R(u) + \alpha \|m - u\|^2. \tag{3}$$

The problem in (3) is divided into two sub-problems, which can be solved using an iterative algorithm:

$$u^{k+1} \leftarrow \min_{u} R(u) + \alpha \|m^k - u\|^2 \tag{4a}$$

$$m^{k+1} \leftarrow \min_{m} \lambda \|Am - y\|^2 + \alpha \|m - u^{k+1}\|^2. \tag{4b}$$

The term in (4a) models the regularization prior which we propose to capture via the learning ability of deep neural networks. The data consistency (DC) constraint is imposed in (4b) and has a closed-form solution. The iterative update steps are given by:

$$u^{k+1} \leftarrow f_{\theta_k}(m^k) \tag{5a}$$

$$m^{k+1} \leftarrow \underbrace{u^{k+1} - \eta A^H (Au^{k+1} - y)}_{\text{DC operation}}, \tag{5b}$$

where f_{θ_k} denotes the k-th sub-network of the reconstruction framework, A^H is the conjugate transpose of A, and $\eta = \frac{\alpha}{\lambda + \alpha}$. The update steps in (5) are unrolled to derive the reconstruction pipeline [38]–[40], by alternating the network reconstruction update and the DC operation. To improve the model performance, we propose an adaptive spatial attention selection module (ASASM) and multi-level dense connections. We devise a deep data consistency block (DCB) to reduce the effect of the bottleneck structure in DC blocks. The reconstruction framework is illustrated in Fig. 2, which we will introduce in detail below.

B. Adaptive Spatial Attention Selection Module (ASASM)

Attention modules, e.g. the transformer and its variants, have been widely used for visual tasks and shown their potential to push performance improvements. To avoid the enormous memory and computational increments, [30] introduces a light-weight attention module, dubbed convolutional block attention module (CBAM). The CBAM module, which comprises a channel gate and a spatial gate, leverages feature statistics, e.g. the average and maximum values, to govern the contribution of different channels and spatial positions.

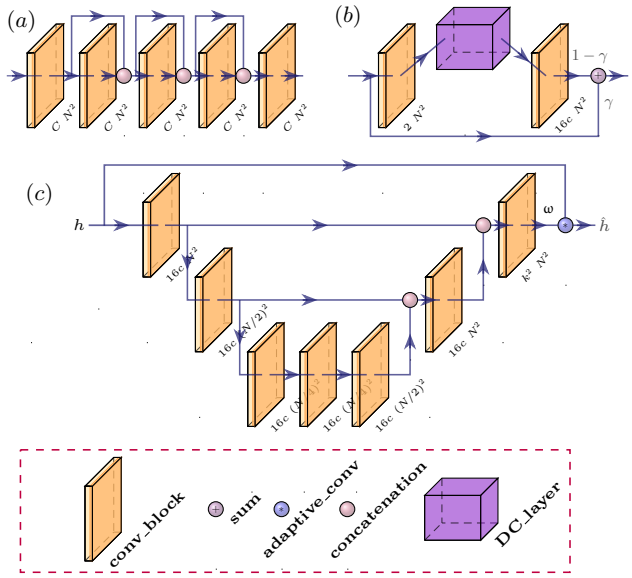

Fig. 1. Illustration of basic modules where C and N^2 indicate the channel size and spatial resolution of output features. a) Densely connected block, b) deep data consistency block (DCB), and c) adaptive spatial attention selection module (ASASM).

However, it can be challenging to provide rich representations using feature statistics. For the purpose of encouraging the network to focus on more important information and achieve notable improvements, we propose an adaptive spatial attention selection module (ASASM), displayed in Fig. 1 (c), which leverages locality-aware spatial attention to endow the learned representations with spatial attention diversities and adaptability. We introduce the architecture of ASASM in the following part of this section.

To overcome the quadratic complexity of conventional attention modules, a spatial gate is introduced in [30] to model the feature statistics to determine the importance of information encoded at each spatial position. However, it can be difficult to handle the complicated relationships among extracted features. We propose to utilize a network structure to predict the spatial attention patterns from the entire features, and perform the spatial attention selection by computing the weighted average of pixels in the neighboring region with adaptively computed kernels. The method in filter adaptive networks [41] computes independent kernels for each pixel and simultaneously introduces channel-wise and spatial variability, i.e. pixels at different channels and spatial positions are averaged with their respective weights to give the output values. The size of the filter prediction is $(C \cdot k^2) \times H \times W$, where C is the number of channels, k^2 represents the kernel size, and $H \times W$ indicates the feature resolution. Consequently, it tremendously increases the computational burden. Different from that, we propose to capture the spatial interdependencies with enlarged receptive fields via an encoding-decoding structure, and subsequently derive the spatial variance in kernel predictions to exploit the diversities of attention patterns and enhance the representation power. As depicted in Fig. 1 (c), the extracted features are passed to a shallow U-net to predict the adaptive kernels which are potentially varying for different

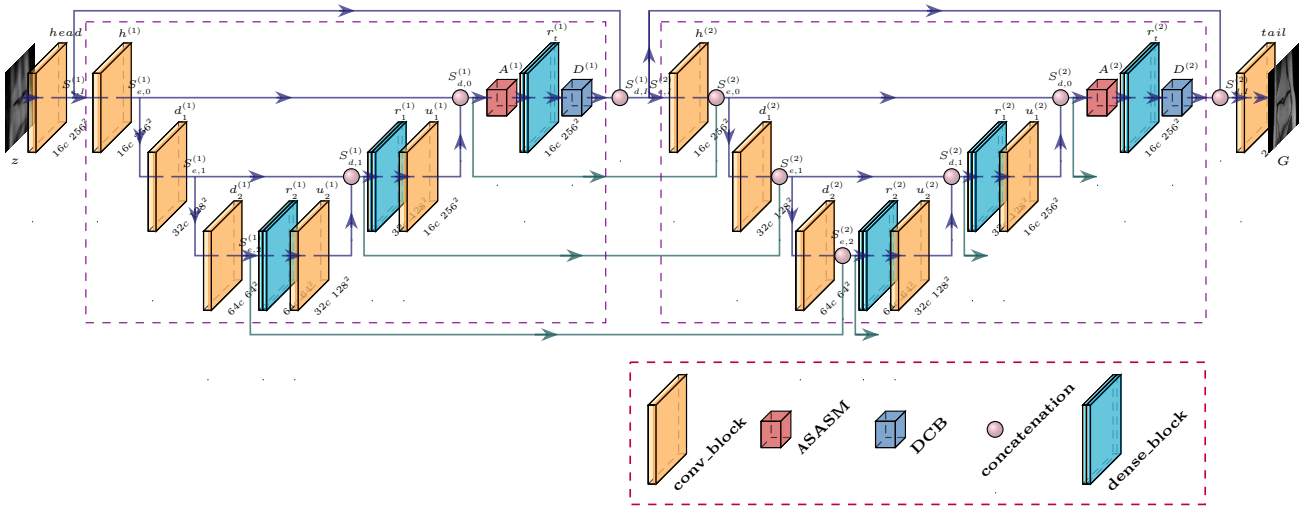

Fig. 2. Overview of the model architecture. The corrupted input is first mapped into feature space by a head block, and then processed by a cascade of U-shaped sub-networks which are densely connected via multi-level shortcuts. Finally, the output features are fused together to give the reconstruction outcome using a tail block. To simplify the illustration, the densely connected design is implemented by concatenating the current output with the collection of all previous predictions, alternating between feature concatenation and skip-connection.

spatial locations. Compared to the reconstruction pipeline, the U-net for kernel prediction has significantly fewer parameters, which can provide a better accuracy-latency trade-off. The size of the predicted filters is reduced to $k^2 \times H \times W$, which greatly decreases the computational and model overhead. The adaptive filtering is applied as follows,

$$\hat{h}(x) = h(x) + \sum_{\delta \in \Delta} \omega(\delta, x) h(x + \delta), \tag{6}$$

where $h(x)$ and $\hat{h}(x)$ denote the input and output feature vectors formed with pixels across the channel axis at spatial position x, Δ represents a regular kernel grid, e.g. $\{-1, 0, 1\}^2$ for $k = 3$, δ is the footprint moving in the filtering sliding window, and $\{\omega(\delta, x)\}_{\delta \in \Delta}$ represents the predicted preferences of the neighboring locations around position x. In order to potentially produce positive attentions and make the gradients manageable, we use a Softmax function to normalize the combination coefficients $\{\omega(\delta, x)\}$, similar to the practice in [42]. We found that in simulations Softmax normalization leads to slightly more stable gradients.

C. Deep Data Consistency Block (DCB)

The incorporation of data consistency (DC) blocks, which "correct" the predicted images with the observed measurements, into deep structures yields considerable improvements in MRI reconstruction, as shown in [33], [38]. Such practice requires to map the predicted features into image domain, which can potentially dispose of important features, due to the feature channel reduction. We use a deep data consistency block (DCB) to correct the predictions and circumvent the bottleneck design via a skip-connected symmetric structure, as displayed in Fig. 1 (b). For simplicity, we assume to follow the single-coil sensing configuration in this section. We first specify two types of data consistency operators, the soft and hard DC operators, and then manifest that our design meets the data consistency requirements and is the extended version of the previous works. The soft DC operator Ω^* mixes the predictions with the observed k-space data on the sampling mask, whereas the hard DC operator Ω replaces the values with the measurements, as shown below:

$$\begin{aligned} \Omega^*(x, \upsilon; y, m) &= F^{-1}(m \odot \frac{F(x) + \upsilon y}{1 + \upsilon} + (1 - m) \odot F(x)) \\ \Omega(x; y, m) &= F^{-1}(m \odot y + (1 - m) \odot F(x)), \end{aligned} \tag{7}$$

where x, υ, and m denote the prediction, a parameter, and the sampling mask, y denotes the measurement given by $y = m \odot F(s)$, F and F^{-1} are the Fourier transform and its inverse, and $\odot$ denotes the element-wise multiplication. Ω^* is utilized in [39], [43] to realize the DC term, and [23], [33] leverage the DC prior [22] to update the synthesized images as shown below,

$$x \leftarrow x - \eta F^{-1}(m \odot F(x) - y), \tag{8}$$

where η is the step size. It can be mathematically proved that the update rules in Ω^* and (8) are equivalent to each other and can be converted to the following form,

$$x \leftarrow \gamma x + (1 - \gamma)\Omega(x; y, m), \tag{9}$$

if the parameters satisfy $\gamma = \frac{1}{1+\upsilon} = 1 - \eta$. We enhance the update rule in (9) by extending it to feature space which potentially contains more information. We show the modified version as follows,

$$h \leftarrow \gamma h + (1 - \gamma) f^*(\Omega(f(h); y, m)), \tag{10}$$

where f and f^* are two convolutional layers used to reduce and expand feature channels, as shown in Fig. 1 (b). By leveraging the skip-connection, we exploit the advantage of residual networks to alleviate the impact of gradient vanishing and facilitate feature propagation.

D. Multi-Level Densely Connected Architecture

Inspired by the success of densely connected layers in image tasks, we introduce the multi-level dense connections to form a novel architecture which potentially refines the learned feature maps with complementary information by "reusing" the synthesized features at multiple levels. To clearly describe the framework, we define a dictionary with items $S_I, S_0, S_1, S_2, ...S_L$, representing the feature maps at different levels of each sub-network including input-output pairs, as illustrated in Fig. 2. L denotes the number of transitional block pairs and implies that in total the model introduces $L+2$ feature levels from image resolution to the topmost feature domain. We utilize the superscript and subscript to separately denote the connection nodes, which can provide convenience to elaborate on the forward computations. Taking $S^{(1)}_{e,\,2}$ for instance, the scripts (1) and $e, 2$ depict that it comes from the second level of the encoding part of the first sub-network.

We demonstrate the proposed architecture in what follows. From simplicity, we only present two sub-networks in the pipeline and select L to be 2, as illustrated in Fig. 2. At each feature level, all previous predictions, not only the preceding one, are collected together and concatenated with the current feature volume to enable the multi-level dense connections. It is noteworthy that the intra- and inter-connections directly start from feature concatenation and the current features are added to the feature collection transmitted from the preceding stage without removing any previous outputs. The densely connected design is therefore applied by alternating between feature concatenation and skip-connection.

The notations of the blocks used in the framework can be found in the annotations in Fig. 2, and we thus evade being necessitated to respectively define them. The initial input to the model is the zero-filled z given by the inverse Fourier transform of the under-sampled measurements. After mapping the two-channel input into feature domain via a head block, we start from the computations of the encoding features at the first sub-network, which are given by gradually extracting information from the previous outputs as follows,

$$\begin{aligned} S^{(1)}_{e,\,I} &\leftarrow head(z) \\ S^{(1)}_{e,\,0} &\leftarrow h^{(1)}(S^{(1)}_{e,\,I}) \\ S^{(1)}_{e,\,1} &\leftarrow d^{(1)}_1(S^{(1)}_{e,\,0}) \\ S^{(1)}_{e,\,2} &\leftarrow d^{(1)}_2(S^{(1)}_{e,\,1}). \end{aligned} \tag{11}$$

Following the design of intra-connections in conventional U-nets, we compute the decoding features as follows,

$$\begin{aligned} S^{(1)}_{d,\,1} &\leftarrow [u^{(1)}_2 \circ r^{(1)}_2(S^{(1)}_{e,\,2}),\, S^{(1)}_{e,\,1}] \\ S^{(1)}_{d,\,0} &\leftarrow [u^{(1)}_1 \circ r^{(1)}_1(S^{(1)}_{d,\,1}),\, S^{(1)}_{e,\,0}] \\ S^{(1)}_{d,\,I} &\leftarrow [D^{(1)} \circ r^{(1)}_t \circ A^{(1)}(S^{(1)}_{d,\,0}),\, S^{(1)}_{e,\,I}], \end{aligned} \tag{12}$$

where $[\,]$ denotes the concatenation of feature maps along the channel axis. We then pass the output dictionary on to the next sub-network. The encoding features are generated with the collections of all preceding outputs at the same levels:

$$\begin{aligned} S^{(2)}_{e,\,I} &\leftarrow S^{(1)}_{d,\,I} \\ S^{(2)}_{e,\,0} &\leftarrow [h^{(2)}(S^{(2)}_{e,\,I}),\, S^{(1)}_{d,\,0}] \\ S^{(2)}_{e,\,1} &\leftarrow [d^{(2)}_1(S^{(2)}_{e,\,0}),\, S^{(1)}_{d,\,1}] \\ S^{(2)}_{e,\,2} &\leftarrow [d^{(2)}_2(S^{(2)}_{e,\,1}),\, S^{(1)}_{e,\,2}]. \end{aligned} \tag{13}$$

Conforming to (12), the decoding procedure is shown below, noting that only intra-connections occur here,

$$\begin{aligned} S^{(2)}_{d,\,1} &\leftarrow [u^{(2)}_2 \circ r^{(2)}_2(S^{(2)}_{e,\,2}),\, S^{(2)}_{e,\,1}] \\ S^{(2)}_{d,\,0} &\leftarrow [u^{(2)}_1 \circ r^{(2)}_1(S^{(2)}_{d,\,1}),\, S^{(2)}_{e,\,0}] \\ S^{(2)}_{d,\,I} &\leftarrow [D^{(2)} \circ r^{(2)}_t \circ A^{(2)}(S^{(2)}_{d,\,0}),\, S^{(2)}_{e,\,I}]. \end{aligned} \tag{14}$$

When we reach the output layer of the final sub-network, e.g. $S^{(2)}_{d,\,I}$ in the current configuration, a tail block is adopted to reduce the channel size and map the synthesized features into images as shown in Fig. 2:

$$G = tail(S^{(2)}_{d,\,I}), \tag{15}$$

where G is the reconstruction result.

E. Final Framework Design

As illustrated in Fig. 2, we utilize two convolutional blocks as the head and tail modules to perform channel expansion and reduction. We deploy the densely connected layers at all decoding feature levels of the U-shaped sub-networks, and embed the ASASM and DCB modules at the top decoding level. The dense connections at multiple levels are adopted to allow intra- and inter- feature propagation. We cascade 4 sub-networks in the reconstruction pipeline and select $c = 1$ for the channel setting.

F. Loss Function Design

In this section, we elaborate on the loss functions used in the training phase. We leverage the high-quality references to supervise the training and enforce the data consistency by comparing the reconstructions with the observed measurements. We encourage sharper and realistic details by training the model in an adversarial diet.

1) Reconstruction Loss: We employ the L_1-norm and structural similarity index metric (SSIM) to quantify the discrepancies between the prediction G and fully-sampled reference s. Compared to L_2-based losses, e.g. the mean squared error (MSE), which potentially produce blurriness in generations, the L_1 loss attempts to mitigate the artifacts caused by L_2. It is suggested in [44] that SSIM-based losses are able to maintain the contrast of high-frequency components and produce visually pleasing results. The reconstruction loss is given as follows,

$$L_{rec} = (1-\alpha)L_1(G, s) + \alpha L^{SSIM}(G, s). \tag{16}$$

where α is set to 0.4.

2) Data Consistency Loss: One challenge of learning a mapping from one domain to the other is to maintain some discriminative information. e.g. semantic attributes, and enforce the synthesized outputs being in agreement with the inputs in terms of salient features. Data consistency terms are proposed in [15], [45] to constrain the solutions within some targeted search space and diminish the mismatching problem. We leverage the consistency loss to encourage the reconstructions to be consistent with the observations, i.e. the under-sampling operation should yield closely matching results. We calculate the consistency loss in k-space as suggested in [46]. An alternative practice is adopted in [18] which applies it in image space. Empirically, we found that they behave very similarly and introduce no fundamental difference. The data consistency loss is calculated as follows,

$$L_{cyc} = \|y - m \odot F(G)\|_1, \tag{17}$$

where the Fourier transform and under-sampling operator are independently applied to each sensitivity coil when dealing with multi-coil data.

3) Adversarial Loss: Generative adversarial networks [37] show great success in producing photo-realistic images. In an adversarial diet, GAN-based models learn the distribution of real data by adopting a discriminator to distinguish generated data from their real counterparts and a generator to fool the discriminator. To address the issues, e.g. the gradient vanishing problem, in Vallina GAN, least squares GAN (LSGAN) is proposed in [47] to prevent the saturation, stabilize the training, and empirically provide faster convergences. The adversarial loss is then given by,

$$\begin{aligned} L_{adv}^{D} &= E[\|D(s) - b\|_2^2] + E[\|D(G) - a\|_2^2] \\ L_{adv}^{G} &= E[\|D(G) - c\|_2^2], \end{aligned} \tag{18}$$

where D denotes the discriminator, E is the expectation, and the hyper-parameters a, b, and c are set to be $a = 0$ and $b = c = 1$ [47]. We adopt the patch-based discriminator in the adversarial diet, also used in CycleGAN [15], which consists of a standard convolutional layer followed by three strided convolutional layers to reduce the spatial size of extracted features.

4) Final Objective: The final objective used in network training is given as follows,

$$L = E_{\{(G,s)\}}[\lambda_{rec} L_{rec} + \lambda_{cyc} L_{cyc} + \lambda_{adv} L_{adv}]. \tag{19}$$

IV. Experiments

In this section, we assess the performance of the proposed framework on both single-coil and multi-coil MRI datasets. The acceleration factors are selected to be $8\times$ and $4\times$ along the phase encoding direction in k-space, as illustrated in Fig. 3, with respectively 4% and 8% central lines preserved and periphery of k-space sampled randomly. The fully sampled k-space data are stored in h5 files and loaded in experiments using the official code package [48]. The under-sampling process is implemented on the k-space raw data and the inverse Fourier transform is applied to map k-space signals to image domain. It is noteworthy that no pre-processing operations, e.g. compressing images which are stored in a database and later used for synthesizing a new k-space and combining multi-coil signals via the root sum of squares (RSS) method to obtain non-negative magnitude images [49], which potentially affect the evaluation of model performance, are used in the simulated encoding process, ensuring that the simulations can be close to real-world applications. We present the comparisons against other state-of-the-art methods. We also display the ablation results to further substantiate the effectiveness of the proposed method. The code of our simulations will be released at https://github.com/JLiu-Edinburgh/HFMRI_Model.

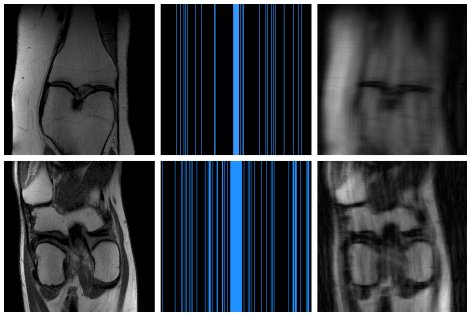

Fig. 3. Illustration of under-sampling patterns at $8\times$ and $4\times$ acceleration. Left) fully sampled, middle) sampling pattern, right) zero-filled.

A. Single-Coil Knee MRI Reconstruction

We use the NYU single-coil knee MRI database [48] to conduct experiments, which contains rich textures and structural details. The data were collected from about 1500 fully sampled MRI scanning cases obtained on 3 and 1.5 Tesla magnets. The raw dataset provides coronal proton density-weighted images. It is distributed as a collection of HDF5 data files, each of which contains the k-space complex-valued data. The dataset was acquired using a coronal proton-density weighting pulse sequence and the following sequence parameters: matrix size 320×320, in-plane resolution $0.5mm \times 0.5mm$, repetition time (TR) ranging between 2200 and 3000 ms, and echo time (TE) between 27 and 34 ms. Due to the limited computational resource, 400 scanning cases are used to train the models and 164 different cases are used in evaluations. The generator takes the zero-filled, which is the inverse Fourier transform of the zero-filled k-space data, as input. We use two channels to handle complex-valued data. The sub-sampling rate is respectively set to 8 and 4, using a fixed random mask. The training parameters are practically set to $\lambda_{rec} = 10$, $\lambda_{cyc} = 0.5$, and $\lambda_{adv} = 0.01$.

We compare the proposed framework with other deep learning methods: MICCAN [43], MoDL [39], FastMRI U-net [48], and ASGAN [32]. MICCAN proposes a deep network with channel-wise attention modules for MRI reconstruction. MoDL combines the representation power of deep neural networks with model-based algorithms to enhance the reconstruction quality. FastMRI U-net reconstructs images from the magnitude maps of the under-sampled signals. ASGAN incorporates the large-field contextual feature integration modules with attention selection in a GAN-based framework to

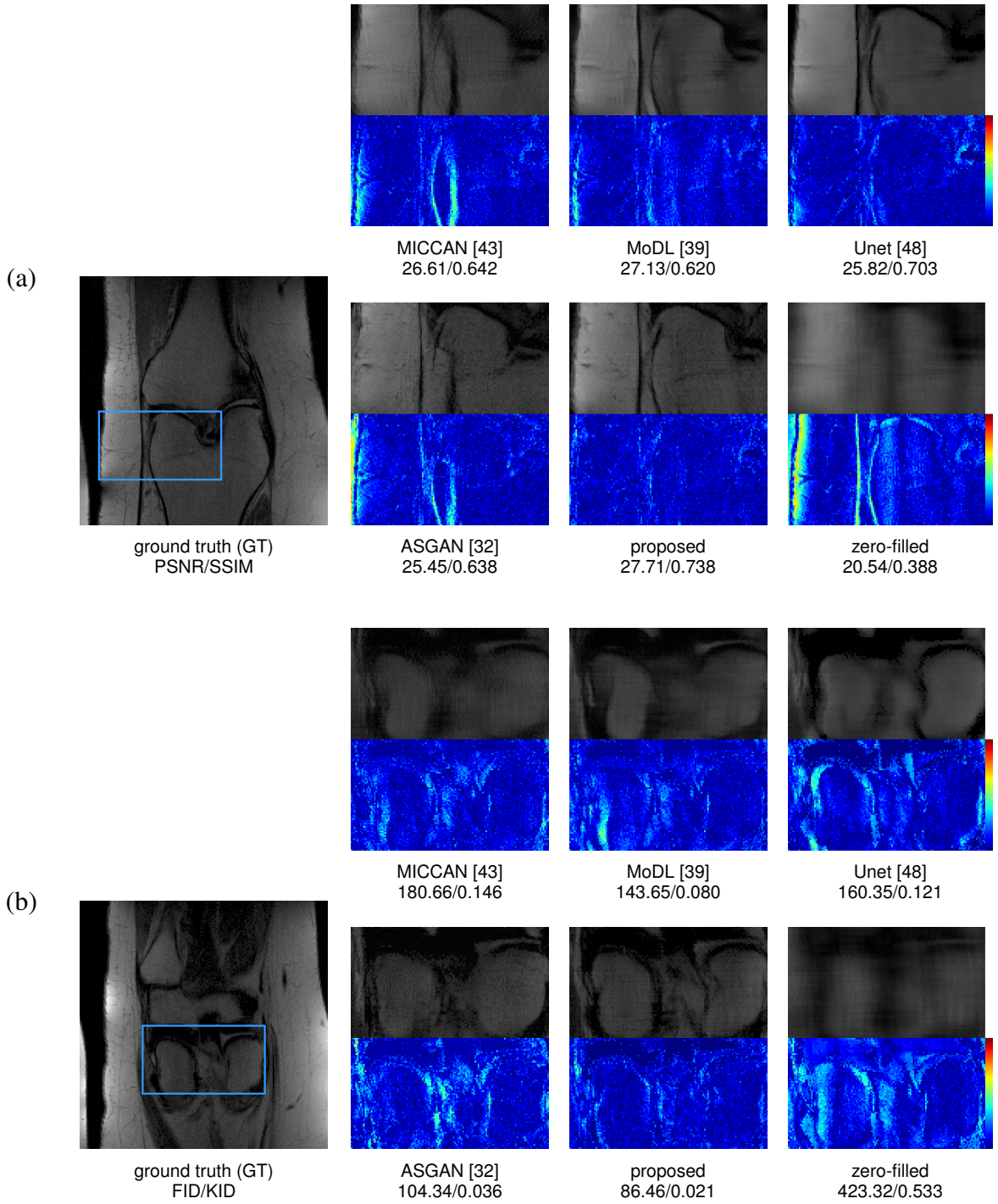

Fig. 4. Comparison results of $8\times$ accelerated knee MRI reconstruction.

generate more local textures and achieve fine-grained MRI reconstructions.

We present the comparison results at $8\times$ and $4\times$ acceleration rates in Fig. 4 and 5. It can be observed that the proposed method produces sharper and more realistic reconstructions, compared with the other approaches. Overall, our framework can preserve salient and informative structures, e.g. Fig. 4 (a), generate more natural and complicated textural features, e.g. Fig. 5 (b), and achieve high-fidelity reconstructions from highly under-sampled measurements. The corresponding residual error maps are displayed in Fig. 4 and 5, where the proposed method yields reconstructions with fewer errors, particularly at a high under-sampling rate.

To evaluate the reconstruction quality using different methods, we adopt PSNR and SSIM as assessment metrics, in which higher values are better. In order to assess the performance of different methods in terms of perceptual quality, the Fréchet inception distance (FID) and the kernel inception

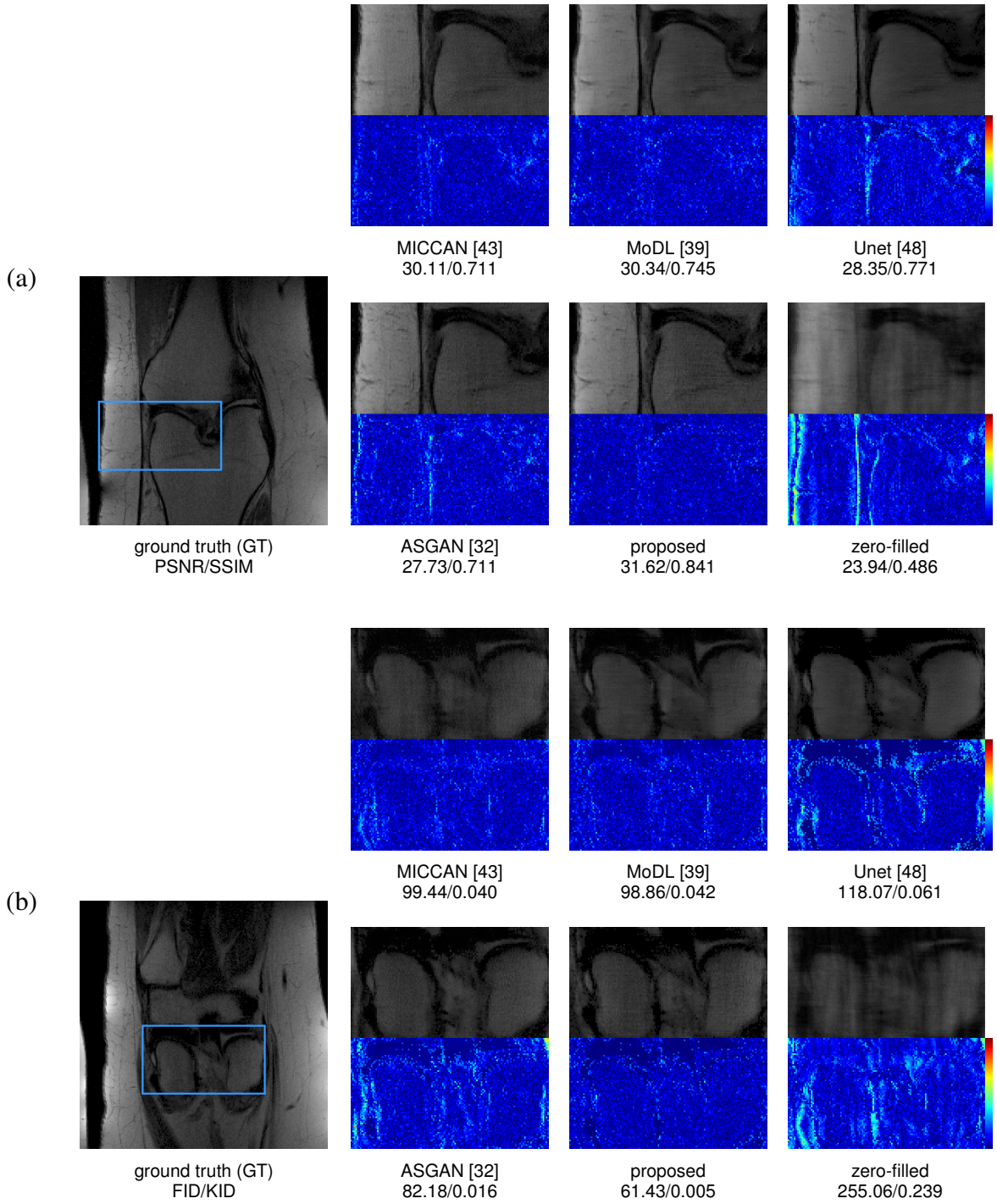

Fig. 5. Comparison results of $4\times$ accelerated knee MRI reconstruction.

distance (KID) [50], in which lower values are preferred, are utilized. The quantitative assessment is performed on the whole imagery with the fully sampled data and p-value statistical test results are also presented. We present the evaluation results in Tab. I. From Tab. I, we can observe that the proposed method consistently produces superior results than the other approaches with remarkable improvements in evaluation metrics, which provides the justification of its effectiveness. We found that MoDL has high PSNR scores. We surmise that it is due to the L_2-based loss used in [39] as the objective function which potentially leads to high PSNR values but fails to correlate well with the structural assessment and the intricate properties of human visual perception [51]. It is noteworthy that the proposed model shows superior performance with fewer parameters than other competing models, and achieves competitive inference speed which potentially enables real-time reconstruction.

TABLE I
QUANTITATIVE EVALUATION ON ACCELERATED KNEE MRI RECONSTRUCTION. 4× ACCELERATED RECONSTRUCTIONS HAVE THE SAME LATENCY (S) AND MODEL SIZE (MB) AS 8×.

	method	PSNR↑	SSIM↑	FID↓	KID↓	p-value	Latency↓	Size↓
8×	proposed	**27.71**(±1.75)	**0.738**(±0.06)	**86.46**(±2.54)	**0.021**(±0.001)	$p < 0.05$	0.028	**4.51**
	ASGAN [32]	25.45(±1.63)	0.638(±0.05)	104.34(±2.43)	0.036(±0.001)	-	0.056	17.0
	Unet [48]	25.82(±1.74)	0.703(±0.08)	160.35(±4.58)	0.121(±0.003)	-	**0.013**	10.5
	MoDL [39]	27.13(±1.68)	0.620(±0.05)	143.65(±3.68)	0.080(±0.003)	-	0.091	22.3
	MICCAN [43]	26.61(±1.77)	0.642(±0.05)	180.66(±5.04)	0.146(±0.004)	-	0.043	10.1
	zero-filled	20.54(±1.29)	0.388(±0.06)	423.32(±7.30)	0.533(±0.005)	-	-	-
4×	proposed	**31.62**(±1.70)	**0.841**(±0.05)	**61.43**(±1.62)	**0.005**(±0.001)	$p < 0.05$	-	-
	ASGAN [32]	27.73(±1.54)	0.711(±0.06)	82.18(±1.85)	0.016(±0.001)	-	-	-
	Unet [48]	28.35(±1.72)	0.771(±0.07)	118.07(±3.93)	0.061(±0.002)	-	-	-
	MoDL [39]	30.34(±1.66)	0.745(±0.05)	98.86(±3.46)	0.042(±0.002)	-	-	-
	MICCAN [43]	30.11(±1.73)	0.711(±0.06)	99.44(±3.82)	0.040(±0.002)	-	-	-
	zero-filled	23.94(±1.45)	0.486(±0.07)	255.06(±4.65)	0.239(±0.003)	-	-	-

B. Multi-Coil Brain MRI Reconstruction

We show that the proposed framework is extensible to multi-coil MRI reconstruction. We train our model on a multi-coil brain MRI dataset from [39]. The dataset was acquired using a 3D T2 CUBE sequence. k-space signals are sensed by multiple coils and the sensitivity maps are provided in the dataset. The matrix dimension of the brain dataset is 256×232 with an isotropic resolution of 1 mm, a field of view (FOV) of $210mm \times 210mm$, and TE=84 ms. It contains 360 scanning samples for training and 164 samples available for test. For multi-coil MRI, the sampling process occurs individually on each coil. The root sum of squares (RSS) method is commonly used to handle the coil correlations by combining the magnitude maps of sensed data. However, the computation of magnitudes disposes of the phase information and can lead to performance degradation. Instead, we simulate the corruption process by projecting the full reconstructions onto each coil sensitivity and computing the sensitivity-weighted average as input to the framework, used also in [38], [39]. The relationships between the sensitivity-weighted average and its coil projections are presented below,

$$\begin{aligned} S_i(s; C_i) &= C_i s \\ s(\{S_i\}; \{C_i\}) &= \frac{\sum_i \bar{C}_i S_i}{\sum_j |C_j|^2}, \end{aligned} \tag{20}$$

where C_i, $\bar{C}_i$, and S_i denote the i-th coil sensitivity map, its conjugate, and the projected signal s with the sensitivity map C_i. Noting that the under-sampling operation is applied to each S_i, the coil sensitivities are individually leveraged in the DCB block to enforce the k-space data consistency, which is formulated as below,

$$h \leftarrow \gamma h + (1-\gamma) f^*(s(\{\Omega(S_i(f(h); C_i); y_i, m)\}, \{C_i\})), \tag{21}$$

where y_i denotes the measurement corresponding to the i-th coil sensitivity. Empirically, we set the training parameters to be $\lambda_{rec} = 30$, $\lambda_{cyc} = 0.5$, and $\lambda_{adv} = 0.005$.

We compare our method with 3 model-based deep learning approaches in $8\times$ and $4\times$ accelerated MRI reconstruction: MoDL [39], ISTA-Net+ [23] which unrolls the shrinkage-thresholding algorithm to resolve the inverse problem, and VS-Net [38] which exploits the iterative process of the variable splitting algorithm in a deep network architecture. The reconstruction results are displayed in Fig. 6 and 7. We can observe that the proposed framework yields shaper edges and clearer details in reconstructed images. From the corresponding residual maps displayed in Fig. 6 and 7, it is shown that the proposed method generates reconstructions with fewer errors. We present the evaluation results in Tab. II. Compared to ISTA-Net+ and VS-Net which outperform MoDL in terms of both visual quality and quantitative evaluations, our model with competitive parameter size and execution speed has improvements in PSNR and SSIM, and posts significant gains in FID and KID, indicating its contributions to the accuracy and perceptual believability of MRI recovery.

C. Ablation Studies on Model Components

We verify the effectiveness of the proposed structure and model components in ablation cases. We first conduct an ablation study to demonstrate the efficacy of the proposed multi-level dense connections. We sequentially cascade the U-shaped sub-networks, as illustrated in Fig. 8 (a), to construct the model denoted by **(A) sequential**. A densely connected variant, dubbed **(B) denseNet**, is derived by only gathering the initial input and all previous outputs of sub-networks together, as displayed in Fig. 8 (b). Compared to **(A)**, it enables dense connections of holistic networks. To balance the model parameter size, we "copy" the encoding and output features of each sub-network for variant **(A)** and **(B)**, which therefore share the parameter overhead of the multi-level connected structure. To substantiate the effectiveness of the proposed model components, we respectively remove the DCB and ASASM modules, which lead to the competing variants **(C) w/o DCB** and **(D) w/o ASASM**. For a fair comparison, the shallow U-net in the ASASM module is preserved to form a residual connection where the output channel size is accordingly adjusted to meet the input size. Similarly, the convolutional layers in DCB are kept for **(C)**. We test the performance of the aforementioned variants on the knee dataset, since knee imagery contains more complicated features and can be suitable to present subtleties and nuances in reconstructions using different methods.

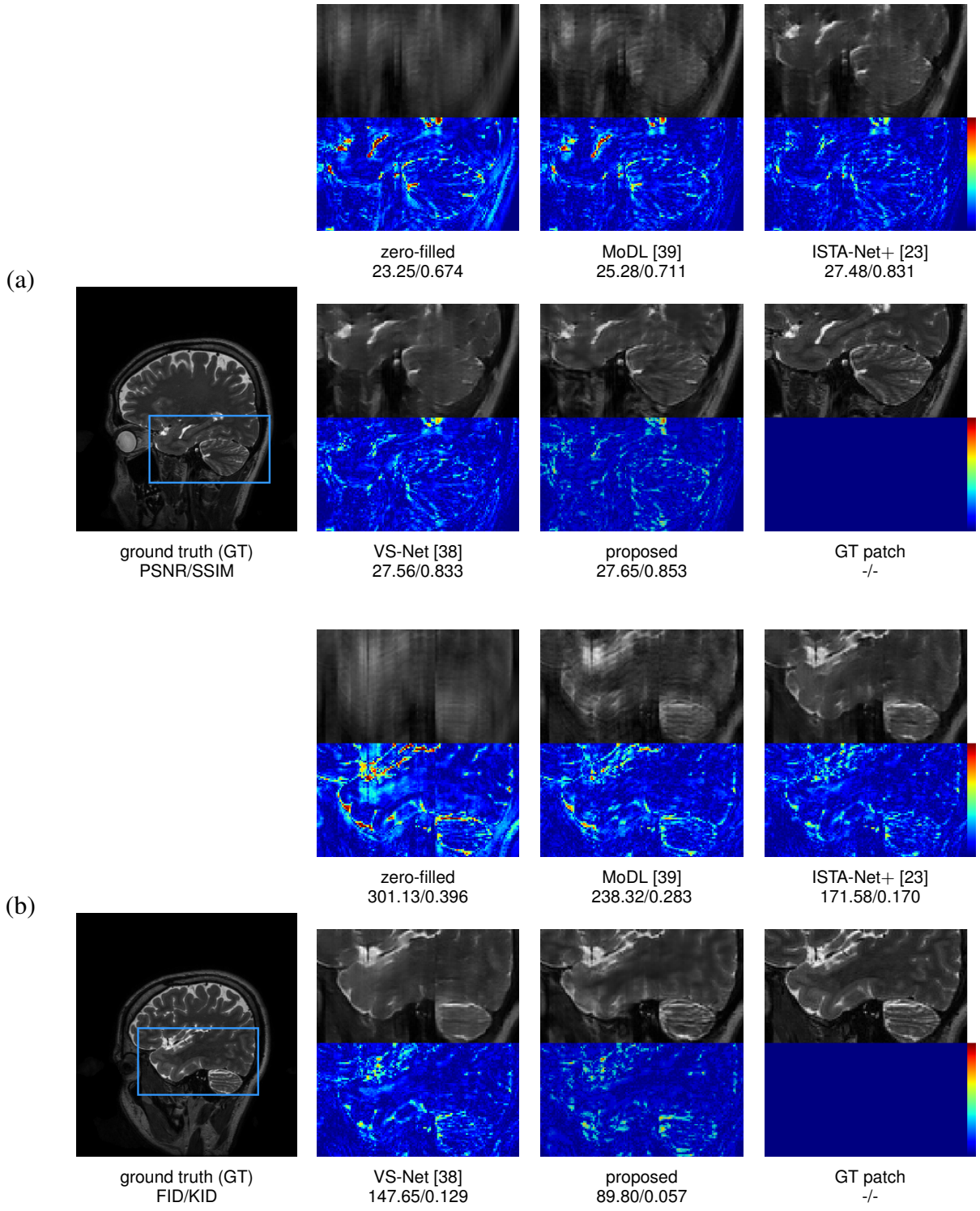

Fig. 6. Comparison results of 8× accelerated brain MRI reconstruction.

We present the evaluation results in Tab. III. It is shown that the proposed multi-level densely connected architecture derives considerably more performance gains than the variant **(B)** with only shortcuts connecting output layers of each sub-network, compared to the sequential model **(A)**. It verifies the usefulness of feature transmission and reuse at multiple levels in pushing performance enhancements. From Tab. III, we observed that both DCB and ASASM can improve the model performance, comparing the variant **(C)** and **(D)** with the proposed model, which confirms their efficacy in pushing reconstruction improvements.

To visualize the performance gains given by the proposed approaches, we display the residual error maps of the ablation cases in Fig. 9. We found that compared to the variant **(A)** the

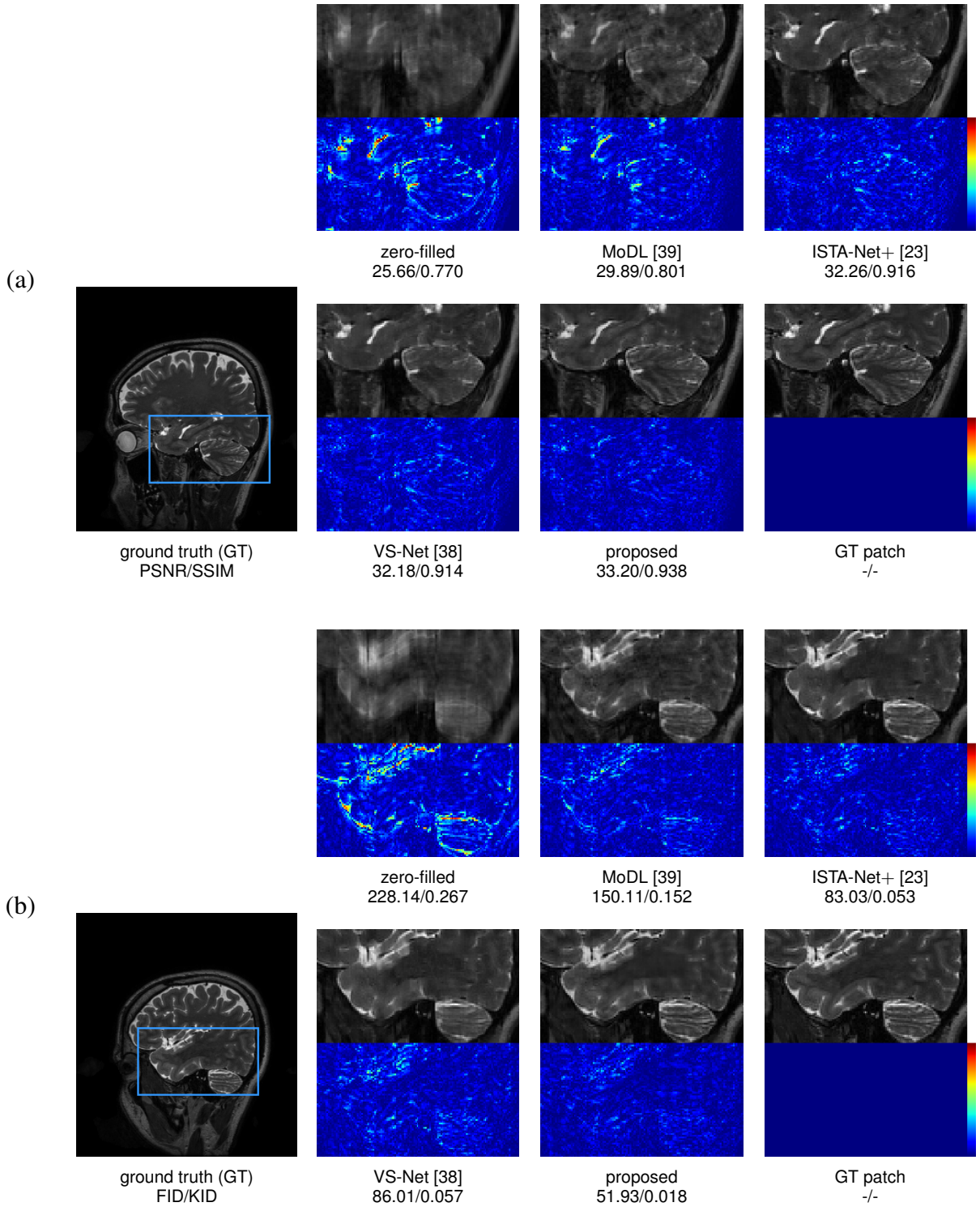

Fig. 7. Comparison results of $4\times$ accelerated brain MRI reconstruction.

proposed multi-level densely connected architecture produces significantly fewer reconstruction errors, while the comparison model **(B)** with dense connections only slightly reduces the errors, which shows the enhanced reconstruction capacity via dense skip-connections at multiple levels. The removal of the DCB module in **(C)** introduces much more errors, demonstrating the necessity of DCB in recovering accurate reconstructions. Compared to **(D)** without the ASASM module, the proposed model produces fewer errors particularly in some local regions, showing the efficacy of ASASM in improving the reconstruction accuracy. In addition, we visualize in Fig. 10 the reconstruction results produced by models with and without the presence of ASASM to further demonstrate its contributions. It can be observed that the model without using

TABLE II
QUANTITATIVE EVALUATION ON ACCELERATED BRAIN MRI RECONSTRUCTION. 4× ACCELERATED RECONSTRUCTIONS HAVE THE SAME LATENCY (S) AND MODEL SIZE (MB) AS 8×.

	method	PSNR↑	SSIM↑	FID↓	KID↓	p-value	Latency↓	Size↓
8×	proposed	**27.65**(±1.66)	**0.853**(±0.05)	**89.80**(±3.85)	**0.057**(±0.001)	$p < 0.05$	**0.031**	4.51
	VS-Net [38]	27.56(±1.69)	0.833(±0.06)	147.65(±4.48)	0.129(±0.002)	-	0.035	4.32
	ISTA-Net+ [23]	27.48(±1.68)	0.831(±0.05)	171.58(±4.57)	0.170(±0.002)	-	0.189	**1.47**
	MoDL [39]	25.28(±1.83)	0.711(±0.04)	238.32(±4.88)	0.283(±0.002)	-	0.094	22.3
	zero-filled	23.25(±2.24)	0.674(±0.07)	301.13(±4.50)	0.396(±0.003)	-	-	-
4×	proposed	**33.20**(±1.10)	**0.938**(±0.02)	**51.93**(±2.33)	**0.018**(±0.001)	$p < 0.05$	-	-
	VS-Net [38]	32.18(±1.19)	0.914(±0.02)	86.01(±3.48)	0.057(±0.002)	-	-	-
	ISTA-Net+ [23]	32.26(±1.18)	0.916(±0.02)	83.03(±2.53)	0.053(±0.002)	-	-	-
	MoDL [39]	29.89(±1.72)	0.801(±0.03)	150.11(±3.52)	0.152(±0.002)	-	-	-
	zero-filled	25.66(±2.06)	0.770(±0.06)	228.14(±3.98)	0.267(±0.002)	-	-	-

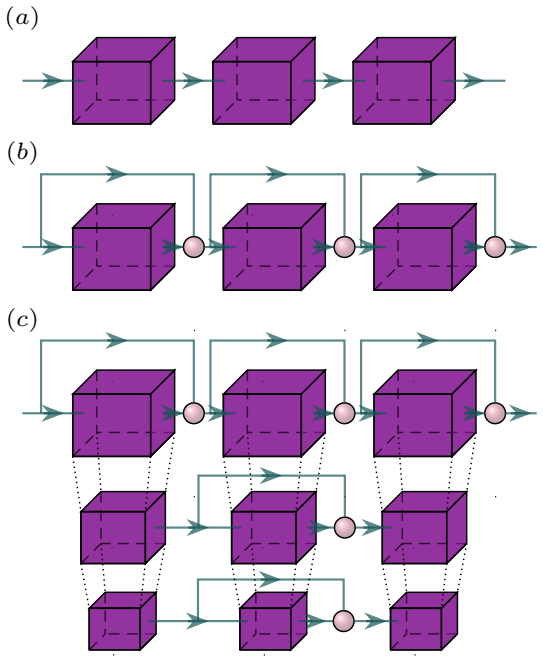

Fig. 8. Illustration of model cascade categories. a) Sequentially connected, b) densely connected, and c) multi-level densely connected. For simplicity, the intra-connections in sub-networks are not displayed in c).

ASASM potentially produces corrupted local structures, e.g. the edge dislocation in Fig. 10, which can be restored when the ASASM module is incorporated into the reconstruction framework. It implies the capacity of the ASASM-integrated model to retrieve contextual information, which is expressed in the form of data-adaptive spatially-varying attention kernels and shows to be beneficial to restoring structural image features. This observation showcases that the devised ASASM can help the network capture local correlations and produce faithful reconstructions. We display the feature maps generated with and without the presence of ASASM in Fig. 11. It can be observed that the feature map produced by the model using ASASM presents clearer and finer structural details, which substantiates the effectiveness of ASASM and demonstrates its capacity to refine the learned feature representation and consequently improve the quality of recovered images.

D. Ablation Studies on Objective Function

We conduct the ablation experiments on the objective function used in our framework. The reconstruction loss measures the differences between the reconstructed image and the ground truth, which ensures the stability and success of the supervised training. The model cannot produce any reasonable reconstruction without it. We therefore compare with two model candidates, dubbed w/o L_{cyc} and w/o L_{adv}, which are respectively trained with either the data consistency loss or the adversarial loss removed. The evaluation results are presented in Tab. IV. We can observe that the candidate trained without L_{cyc} shows slightly degraded performance in terms of all evaluation metrics, which indicates the usefulness of the data consistency constraint adopted in the training objective. Compared with the proposed model, candidate w/o L_{adv} produces higher PSNR and SSIM scores by a small margin, while showing considerably increased perceptual distances in terms of FID and KID. To strike a better trade-off between reconstruction accuracy and image quality, the combination of the three objective losses is used in our experiments.

E. Comparison Study on Data Consistency Blocks

To further validate the efficacy of the DCB module, we compare the proposed method with a conventional data consistency operation by removing the feature skip-connection in DCB and implementing the general update rule in (9). The predicted features are therefore updated as follows,

$$h \leftarrow f^*(\gamma f(h) + (1-\gamma)\Omega(f(h); y, m)). \tag{22}$$

We denote by **w/o shortcut** the resulting model and present the comparison results in Tab. V. We discovered that the conventional DC operation shows inferior performance in all evaluation metrics, e.g. it drops by 0.5dB in PSNR and reaches 93.51 in FID. It indicates the negative impact of the bottleneck design in DC operators, and confirms that the proposed DCB can alleviate this issue by easier feature transmission and boost reconstruction performance.

F. Reconstruction Results with Random Noises

We conducted experiments with single-coil and multi-coil datasets and two types of anatomical structures: knee and

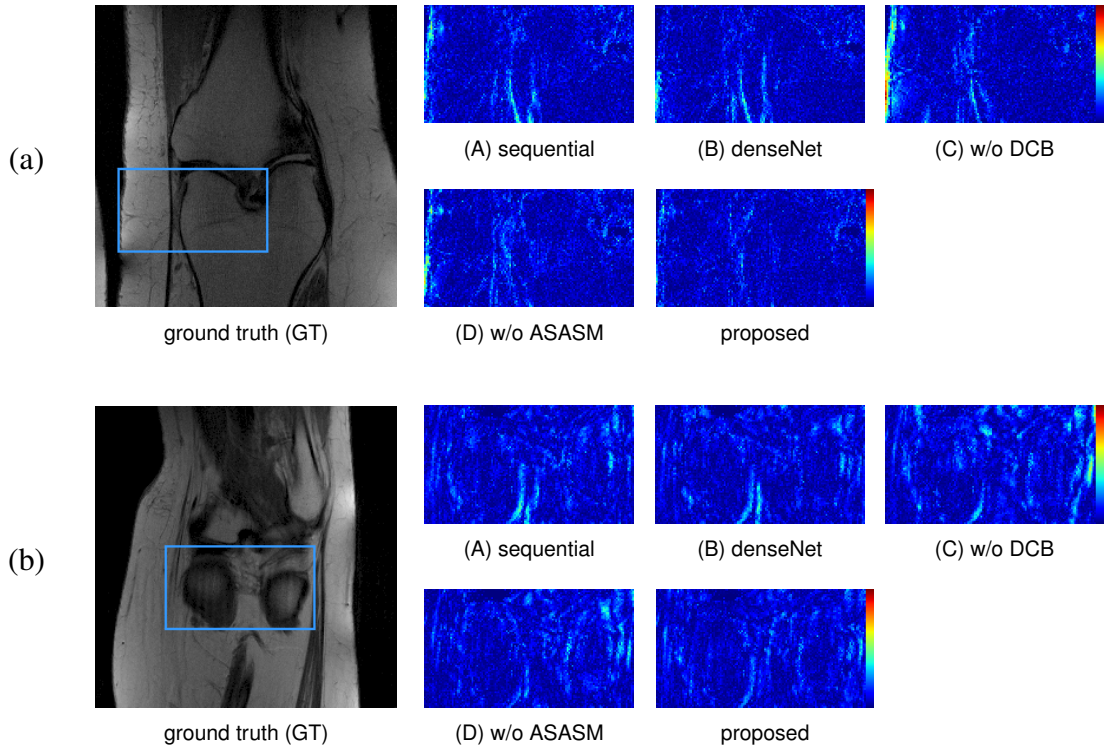

Fig. 9. Ablation residual maps ($2\times$ amplified) of $8\times$ accelerated knee reconstruction.

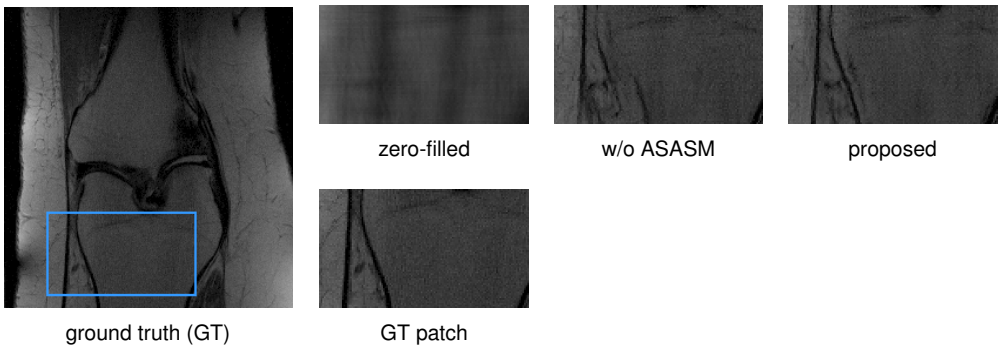

Fig. 10. Ablation results with and without the presence of ASASM at $8\times$ acceleration.

TABLE III
ABLATION STUDIES ON THE PROPOSED MODEL COMPONENTS AND STRUCTURE USING KNEE DATASET

method	PSNR↑	SSIM↑	FID↓	KID↓	p-value
(A) sequential	27.25(±1.69)-*0.46*	0.725(±0.06)-*0.013*	91.33(±2.74)+*4.87*	0.024(±0.001)+*0.003*	-
(B) denseNet	27.49(±1.73)-*0.22*	0.731(±0.06)-*0.007*	89.91(±2.48)+*3.45*	0.023(±0.001)+*0.002*	-
(C) w/o DCB	26.13(±1.70)-*1.58*	0.705(±0.06)-*0.033*	95.50(±3.19)+*9.04*	0.029(±0.001)+*0.008*	-
(D) w/o ASASM	27.15(±1.80)-*0.56*	0.726(±0.07)-*0.012*	94.23(±2.60)+*7.77*	0.027(±0.001)+*0.006*	-
proposed	**27.71**(±1.75)	**0.738**(±0.06)	**86.46**(±2.54)	**0.021**(±0.001)	$p < 0.05$

TABLE IV
ABLATION STUDY ON OBJECTIVE FUNCTION USING KNEE DATASET

method	PSNR↑	SSIM↑	FID↓	KID↓	p-value
proposed	**27.71**(±1.75)	**0.738**(±0.06)	**86.46**(±2.54)	**0.021**(±0.001)	$p < 0.05$
w/o L_{cyc}	27.57(±1.81)-*0.14*	0.734(±0.06)-*0.004*	88.06(±2.66)+*1.6*	0.023(±0.001)+*0.002*	-
w/o L_{adv}	27.88(±1.83)+*0.17*	0.750(±0.06)+*0.012*	106.96(±2.79)+*20.5*	0.049(±0.001)+*0.028*	-

TABLE V
COMPARISON STUDY ON DATA CONSISTENCY BLOCKS USING KNEE DATASET

method	DC	PSNR↑	SSIM↑	FID↓	KID↓	p-value
proposed	DCB	**27.71**(±1.75)	**0.738**(±0.06)	**86.46**(±2.54)	**0.021**(±0.001)	$p < 0.05$
w/o shortcut	general	27.21(±1.76)-*0.5*	0.730(±0.06)-*0.008*	93.51(±3.06)+*7.05*	0.026(±0.001)+*0.005*	-

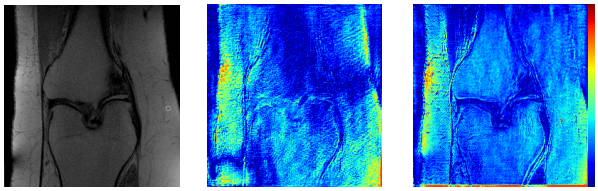

Fig. 11. Feature map visualization with and without the presence of ASASM at $8\times$ acceleration.

brain. It was demonstrated that the proposed model can generalize well on different datasets under both single-coil and multi-coil configurations. In real-world applications, the observed measurements can be unpredictably corrupted by noises, possibly leading to degraded reconstruction performance. From the results displayed in Fig. 12, where random Gaussian noises with different deviations are added to the measurements in inference, we found that the noises increase the errors appearing in the residual maps. Meanwhile, the recovered images present less faithful local textures, e.g. the result with $\sigma = 0.025$, compared to the reconstruction without being corrupted by noises. Those observations have implications that adding noises to k-space data potentially imposes a more challenging problem, which can be formed as a mixture of image de-noising and reconstruction tasks and requires improving the model robustness capacity. It is an interesting research direction to incorporate the de-noising task into MRI reconstruction with the object of enhancing the robustness and generalization capacity of reconstruction frameworks.

V. DISCUSSIONS

The simulation results show that the proposed optimization-inspired MRI reconstruction framework outperforms other state-of-the-art methods qualitatively and quantitatively on both knee and brain datasets visually and in evaluation metrics. In ablation results, we found that the multi-level densely connected structure introduces much more improvements as compared to the model only adopting dense connections between holistic sub-networks, which showcases the benefit of information transmission and reuse at multiple feature levels. Compared with the other proposed modules, the DCB delivers prominent reconstruction improvements and removing it leads to higher errors as shown in Fig. 9, which is due to the two advantages of the proposed data consistency method. Firstly, the DCB explicitly enforces the k-space data constraint and corrects the intermediate predictions with the observed measurements when iterating the reconstruction step, which consequently encourages more faithful results to the target references. Secondly, the residual shortcut used in the DCB in feature space avoids the bottleneck design in traditional data consistency blocks and shows its ability to enhance the network performance, as demonstrated in Sect. IV-E. The proposed ASASM module endows the network with the capacity to perform data-adaptive kernel prediction and encode the captured contextual information in spatially-varying attention patterns. Although the obtained improvements are not as significant as adopting the DCB, it is demonstrated in Fig. 10 that employing the ASASM module can alleviate the corrupted image details and better preserve structures. Those observations suggest the necessity of the proposed network components and imply a synergistic effect in attaining performance advantages.

In experiments, retrospective under-sampling strategies were adopted to simulate the accelerated data acquisition procedure, since the datasets only provide fully sampled k-space signals to demonstrate the effectiveness of devised reconstruction methods and prospectively acquired data are not available. In such cases, reconstruction models are trained with the retrospectively under-sampled phase encodes, and a consistent acquisition pipeline with a prospective sampling approach can be provided with a subset of phase encodes pre-determined and acquired [39]. However, in some cases such as T_2-weighted echo train imaging sequences, the retrospectively under-sampled data can have a higher signal-to-noise ratio (SNR) than that of prospectively acquired data, due to the difference in T_2-weighting between the two sampling methods, which can lead to enhanced correspondence to the ground truth and over-estimated evaluation results [52]. Meanwhile, prospectively sampled datasets can be utilized to qualitatively evaluate the generalizability and validity of reconstruction methods and more faithfully demonstrate the reconstruction performance in real-world applications [53], [54]. In light of these considerations, it can be a crucial research direction to explore the potential of reconstruction models which take into consideration both retrospective and prospective sampling methods, aiming to alleviate the issues caused by the discrepancies between the two under-sampling schemes and enhance the feasibility of the proposed methods in clinical applications.

In experiments, we respectively trained the models on two MRI datasets. More training samples can potentially enable us to train a higher performance model with increased complexity. However, it incredibly requires more training costs and time. Meanwhile, the data acquisition of MR images can be more expensive than that of natural images, which sometimes prevents us from establishing a model on very large-scale MRI datasets. As opposed to end-to-end network models which directly learn the mapping from under-sampled signals to clean images, incorporating domain-specific knowledge into

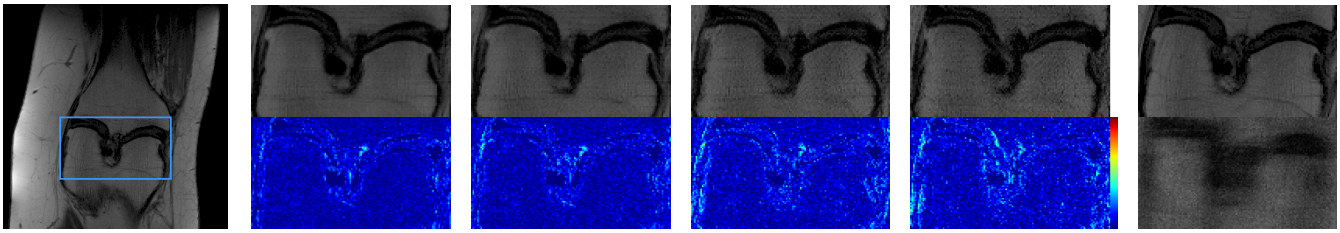

Fig. 12. Reconstruction results with random Gaussian noises added to the measurements in inference.

reconstruction frameworks, e.g. in the form of the unrolling of regularized optimization algorithms, enables us to develop reconstruction models with fewer samples.

As an alternative to Cartesian sampling patterns, non-Cartesian sampling strategies, e.g. radial sampling scheme, are also common practice. However, it demands the use of the non-uniform fast Fourier transform and gridding [55], [56], which can propagate interpolation errors and unfavorably affect the reconstruction performance [34]. We therefore focus on Cartesian sampling methods in this paper and leave it for future work to explore other types of non-Cartesian sampling patterns.

In quantitative evaluations, the FID and KID were adopted as the perceptual metrics. They exploit deep neural networks to measure the differences between the reconstructed images and the ground truths in feature spaces learned from large-scale image datasets, which are potentially more consistent with the human perceptual judgement system. Considering the specific anatomical structures in MRI modality, it is worth exploring and developing domain-specific evaluation methods to gauge the visual quality of MRI reconstructions.

VI. Conclusion

In this paper, we introduce a novel framework to achieve high-fidelity MRI reconstruction. We propose two neural network structures to derive adaptive spatial attention and encourage data consistency. In experiments, we demonstrate that on top of a novel multi-level densely connected architecture the proposed framework can notably promote excellence in recovering highly corrupted MR images. For the future works, it is worth investigating the role of data consistency terms in reconstruction tasks. Another interesting research direction would be to extend the model to dynamic MRI reconstruction.

References

[1] M. Griswold, P. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic resonance in medicine*, vol. 47, no. 6, p. 1202—1210, June 2002.

[2] M. Fair, P. Gatehouse, E. DiBella, and D. Firmin, "A review of 3D first-pass, whole-heart, myocardial perfusion cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 17, 2015.

[3] M. Hong, Y. Yu, H. Wang, F. Liu, and S. Crozier, "Compressed sensing MRI with singular value decomposition-based sparsity basis," *Physics in Medicine and Biology*, vol. 56, no. 19, pp. 6311–6325, Sep 2021.

[4] S. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.

[5] S. Roohi, D. Zonoobi, A. Kassim, and J. Jaremko, "Multi-dimensional low rank plus sparse decomposition for reconstruction of under-sampled dynamic MRI," *Pattern Recognition*, vol. 63, pp. 667–679, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320316302965

[6] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1028–1041, 2011.

[7] S. Anuroop, Z. Jure, M. Tullie, Z. Lawrence, D. Aaron, and D. K.S., "GrappaNet: combining parallel imaging with deep learning for multi-coil MRI reconstruction," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14 303–14 310, 2020.

[8] D. Lee, J. Yoo, S. Tak, and J. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Trans. Biomed. Eng*, vol. 65, no. 9, pp. 1985–1995, 2018.

[9] K. Tezcan, C. Baumgartner, R. Luechinger, K. Pruessmann, and E. Konukoglu, "MR image reconstruction using deep density priors," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1633–1642, 2019.

[10] I. Goodfellow, A. Pouget, M. Mirza, B. Xu, F. Warde, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680, 2014.

[11] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, 2018.

[12] D. Narnhofer, K. Hammernik, F. Knoll, and T. Pock, "Inverse GANs for accelerated MRI reconstruction," *Wavelets and Sparsity XVIII*, vol. 11138, p. 111381A, Sep. 2019.

[13] S. Bhadra, W. Zhou, and M. Anastasio, "Medical image reconstruction with image-adaptive priors learned by use of generative adversarial networks," *Medical Imaging 2020: Physics of Medical Imaging*, vol. 11312, pp. 206 – 213, 2020.

[14] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 167–179, 2019.

[15] J. Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, 2017.

[16] K. Armanious, C. Jiang, S. Abdulatif, T. Küstner, S. Gatidis, and B. Yang, "Unsupervised medical image translation Using Cycle-MedGAN," *27th European Signal Processing Conference, EUSIPCO 2019, A Coruña, Spain, September 2-6, 2019*, pp. 1–5, 2019.

[17] L. Gatys, A. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2414–2423, June 2016.

[18] B. Sim, G. Oh, S. Lim, C. Jung, and J. Ye, "Optimal transport driven CycleGAN for unsupervised learning in inverse problems," *SIAM Journal on Imaging Sciences*, vol. 13, no. 4, pp. 2281–2306, 2020.

[19] C. Zhang, Y. Liu, F. Shang, Y. Li, and H. Liu, "A novel learned primal-dual network for image compressive sensing," *IEEE Access*, vol. 9, pp. 26 041–26 050, 2021.

[20] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[21] Y. Liu, Q. Liu, M. Zhang, Q. Yang, S. Wang, and D. Liang, "IFR-Net: Iterative feature refinement network for compressed sensing MRI," *IEEE Trans. Comput. Imag*, vol. 6, 2020.

[22] N. Pezzotti, S. Yousefi, M. Elmahdy, J. van Gemert, C. Schülke, M. Doneva, T. Nielsen, S. Kastryulin, B. F. Lelieveldt, M. P. van Osch, E. de Weerdt, and M. Staring, "An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction," *arXiv e-prints*, p. arXiv:2004.07339, Apr. 2020.

[23] J. Zhang and B. Ghanem, "ISTA-Net: iterative shrinkage-thresholding algorithm inspired deep network for image compressive sensing," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1828–1837, 2018.

[24] R. Liu, Y. Zhang, S. Cheng, Z. Luo, and X. Fan, "A deep framework assembling principled modules for CS-MRI: unrolling perspective, convergence behaviors, and practical modeling," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4150–4163, 2020.

[25] R. Ahmad, C. Bouman, T. Buzzard, S. Chan, S. Liu, E. Reehorst, and P. Schniter, "Plug-and-play methods for magnetic resonance imaging: using denoisers for image recovery," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 105–116, Jan. 2020.

[26] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, vol. 97, pp. 7354–7363, 2019.

[27] Z. Yuan, M. Jiang, Y. Wang, B. Wei, Y. Li, P. Wang, W. Menpes-Smith, Z. Niu, and G. Yang, "SARA-GAN: self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing MRI reconstruction," *Frontiers in Neuroinformatics*, vol. 14, pp. 1–12, 2020.

[28] Y. Wu, Y. Ma, J. Liu, J. Du, and L. Xing, "Self-attention convolutional neural network for improved MR image reconstruction," *Information Sciences*, vol. 490, pp. 317–328, 2019.

[29] T. Du, H. Zhang, Y. Li, H. Song, and Y. Fan, "Adaptive convolutional neural networks for k-space data interpolation in fast magnetic resonance imaging," *CoRR*, vol. abs/2006.01385, 2020.

[30] S. Woo, J. Park, J. Lee, and I. Kweon, "CBAM: convolutional block attention module," *CoRR*, vol. abs/1807.06521, 2018.

[31] Y. Guo, C. Wang, H. Zhang, and G. Yang, "Deep attentive Wasserstein generative adversarial networks for MRI reconstruction with recurrent context-awareness," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 167–177, 2020.

[32] J. Liu and M. Yaghoobi, "Fine-grained MRI reconstruction using attentive selection generative adversarial networks," *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1155–1159, 2021.

[33] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. Zitnick, N. Yakubova, F. Knoll, and P. Johnson, "End-to-end variational networks for accelerated MRI reconstruction," *ArXiv*, vol. abs/2004.06688, 2020.

[34] J. Schlemper, J. Caballero, J. Hajnal, A. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2018.

[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[36] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[37] K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. Loy, Y. Qiao, and X. Tang, "ESRGAN: enhanced super-resolution generative adversarial networks," *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018.

[38] J. Duan, J. Schlemper, C. Qin, C. Ouyang, W. Bai, C. Biffi, G. Bello, B. Statton, D. Regan, and D. Rueckert, "VS-Net: variable splitting network for accelerated parallel MRI reconstruction," *Springer International Publishing*, pp. 713–722, 2019.

[39] H. Aggarwal, M. Mani, and M. Jacob, "MoDL: model-based deep learning architecture for inverse problems," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394–405, 2019.

[40] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep magnetic resonance image reconstruction: inverse problems meet neural networks," *IEEE Signal Process Mag*, vol. 37, pp. 141–151, 2020.

[41] S. Zhou, J. Zhang, J. Pan, H. Xie, W. Zuo, and J. Ren, "Spatio-Temporal Filter Adaptive Network for Video Deblurring," *Proceedings of the IEEE International Conference on Computer Vision*, 2019.

[42] S. Bako, T. Vogels, B. McWilliams, M. Meyer, J. Novák, A. Harvill, P. Sen, T. DeRose, and F. Rousselle, "Kernel-predicting convolutional networks for denoising Monte Carlo renderings," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, July 2017.

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

[44] P. Deora, B. Vasudeva, S. Bhattacharya, and M. P.P., "Structure preserving compressive sensing MRI reconstruction using generative adversarial networks," *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[45] J. Zhu, R. Zhang, D. Pathak, T. Darrell, A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," *Advances in Neural Information Processing Systems*, 2017.

[46] T. Quan, D. Nguyen, and W. Jeong, "Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1488–1497, 2018.

[47] X. Mao, Q. Li, H. Xie, R. Lau, Z. Wang, and S. Smolley, "Least squares generative adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, 2017.

[48] J. Zbontar, F. Knoll, A. Sriram, M. Muckley, M. Bruno, A. Defazio, M. Parente, K. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. Zitnick, M. Recht, D. Sodickson, and Y. Lui, "FastMRI: an open dataset and benchmarks for accelerated MRI," *CoRR*, vol. abs/1811.08839, 2018.

[49] E. Shimron, J. Tamir, K. Wang, and M. Lustig, "Implicit data crimes: machine learning bias arising from misuse of public data," *Proceedings of the National Academy of Sciences*, vol. 119, no. 13, p. e2117203119, 2022.

[50] M. Bińkowski, D. Sutherland, M. Arbel, and A. Gretton, "Demystifying MMD GANs," *International Conference on Learning Representations*, 2018.

[51] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[52] D. McClymont, I. Teh, H. Whittington, V. Grau, and J. Schneider, "Prospective acceleration of diffusion tensor imaging with compressed sensing using adaptive dictionaries," *Magnetic Resonance in Medicine*, vol. 76, no. 1, pp. 248–258, 2016.

[53] Z. Ke, W. Huang, Z. Cui, J. Cheng, S. Jia, H. Wang, X. Liu, H. Zheng, L. Ying, Y. Zhu, and D. Liang, "Learned low-rank priors in dynamic MR imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3698–3710, 2021.

[54] W. Huang, Z. Ke, Z. Cui, J. Cheng, Z. Qiu, S. Jia, L. Ying, Y. Zhu, and D. Liang, "Deep low-rank plus sparse network for dynamic MR imaging," *Medical Image Analysis*, vol. 73, p. 102190, 2021.

[55] J. Fessler and B. Sutton, "Nonuniform fast Fourier transforms using min-max interpolation," *IEEE Trans. Signal Process.*, vol. 51, 2003.

[56] J. Fessler, "On NUFFT-based gridding for non-Cartesian MRI," *Journal of magnetic resonance*, vol. 188, no. 2, 2007.