DONet: Dual-Octave Network for Fast MR Image Reconstruction

Chun-Mei Feng, Zhanyuan Yang, Huazhu Fu, Senior Member, IEEE, Yong Xu, Senior Member, IEEE, Jian Yang, Member, IEEE, and Ling Shao, Fellow, IEEE

Abstract—Magnetic resonance (MR) image acquisition is an inherently prolonged process, whose acceleration has long been the subject of research. This is commonly achieved by obtaining multiple undersampled images, simultaneously, through parallel imaging. In this article, we propose the dual-octave network (DONet), which is capable of learning multiscale spatial-frequency features from both the real and imaginary components of MR data, for parallel fast MR image reconstruction. More specifically, our DONet consists of a series of dual-octave convolutions (Dual-OctConvs), which are connected in a dense manner for better reuse of features. In each Dual-OctConv, the input feature maps and convolutional kernels are first split into two components (i.e., real and imaginary) and then divided into four groups according to their spatial frequencies. Then, our Dual-OctConv conducts intragroup information updating and intergroup information exchange to aggregate the contextual information across different groups. Our framework provides three appealing benefits: 1) it encourages information interaction and fusion between the real and imaginary components at various spatial frequencies to achieve richer representational capacity; 2) the dense connections between the real and imaginary groups in each Dual-OctConv make the propagation of features more efficient by feature reuse; and 3) DONet enlarges the receptive field by learning multiple spatial-frequency features of both the real and imaginary components. Extensive experiments on two popular datasets (i.e., clinical knee and fastMRI), under different undersampling patterns and acceleration factors, demonstrate the superiority of our model in accelerated parallel MR image reconstruction.

Index Terms—Complex-valued data, feature fusion, image reconstruction, magnetic resonance (MR) imaging.

I. INTRODUCTION

MAGNETIC resonance (MR) imaging has become increasingly popular in radiology and medicine over the past decade due to its advantages in being nonradiative, having a high spatial resolution, and providing superior soft tissue contrast [1]. However, a major limitation of MR imaging is that it requires a much longer acquisition time than other imaging techniques, e.g., computed tomography (CT), X-ray, and ultrasound [2], [35]. In addition, it is impossible to keep patients in the scanner for a long time and obtain clean data without motion artifacts. Therefore, accelerating MR reconstruction has become an urgent research problem since it can, in turn, greatly accelerate MR imaging. Recently, significant efforts have been devoted to this task, which is typically achieved by reconstructing the desired full images from undersampled measured data [3], [4].

Compressed sensing (CS) has achieved significant progress in fast MR imaging because sub-Nyquist sampling can significantly reduce the acquisition time by skipping some phase information. Specifically, CS-based methods overcome aliasing artifacts caused by the violation of the Shannon–Nyquist sampling theorem by introducing additional prior knowledge of the images. CS-based methods employ sparse coding in the transformed domain (e.g., undersampled k-space data) to naturally compress the MR image through a discrete cosine transform (DCT) [5], [6], discrete Fourier transform (DFT) [7], [8], discrete wavelet transform (DWT) [9]–[11], or dictionary learning [12]. However, these traditional methods only exploit prior information of the images-to-be-reconstructed or involve very few reference images. For a large number of medical images acquired in clinical practice, traditional CS-based methods do not explore the potential regularities within them. Because CS-based methods require iterative optimization to find the optimal value, even if the CS involves only a small number of reference images, they still suffer from heavy computational overhead.

Recently, parallel MR imaging has been considered as one of the most important achievements in accelerated MR imaging [13]–[16]. Multicoil data for parallel imaging...
is composed of multiple physical receivers simultaneously recorded from different perspectives. Parallel MR imaging reconstructs the data points using the coil sensitivity profiles across multiple channels, from a small amount of $k$-space data. At present, parallel imaging is the default option for many scanning protocols. Most studies (e.g., SENSE [17], GRAPPA [18], [19], and SPIRiT [20]) take advantage of spatial sensitivity and gradient coding to reduce the amount of data required for reconstruction, thereby shortening the imaging time. However, a higher acceleration rate may introduce aliased artifacts and substantially reduce the signal-to-noise ratio (SNR) in clinical practice.

More recently, with the renaissance of deep neural networks, deep learning techniques, especially convolutional neural networks (CNNs) [21], [22], have been widely used for parallel MR imaging [23], [24]. Since models are trained offline over large-scale data, only a few extra online samples are required for reconstruction. The model-based unrolling methods [25], [26] combine mathematical structures (e.g., variational inference and CS) with deep learning for fast MR image reconstruction. Moreover, extensive approaches [2], [27]–[30] introduce end-to-end learnable models to remove the aliasing artifacts from images that are reconstructed from undersampled multicoil $k$-space data. The mapping between a zero-filled $k$-space and fully sampled MR image is automatically learned by CNNs, requiring no subproblem division.

Most of the above approaches directly borrow vanilla convolutions used in standard CNNs for $k$-space data in MR image reconstruction. However, vanilla convolutions are designed for real-valued natural images and cannot deal with complex-valued inputs. To solve this, early studies [31] simply discarded the imaginary part or processed the real and imaginary parts independently for real-valued convolutions [see Fig. 1(a)]. To avoid information loss, the complex convolution [32] was recently proposed to process complex-valued inputs and encourage information exchange between real and imaginary values [see Fig. 1(b)]. Though impressive, existing complex convolution operations ignore the intrinsic multifrequency property of MR images, leading to limited single-scale contextual information and high spatial redundancy in the final representations.

To address these limitations, we take a further step toward exploring multifrequency representation learning in parallel MR image reconstruction [see Fig. 1(c)]. Specifically, we propose a novel dual-octave network (DONet) that contains a series of dual-octave convolutions (Dual-OctConvs), enabling our model to learn multifrequency representations of multicoil MR images [33]. The dense connections in the Dual-OctConv enable the feature propagation of the high- and low-frequency components of both the real and imaginary data. Unlike complex convolutions, our DONet processes the real (or imaginary) part of MR image features by factorizing it into high- and low-frequency components. The low-frequency component shares information across neighboring locations and can, thus, be efficiently processed in low resolution to enlarge the receptive field and reduce the spatial redundancy. Finally, we combine the features of the real and imaginary parts for reconstruction. Benefiting from the Dual-OctConv, our network has a more powerful capability in multiscale representation learning and can, thus, better capture soft tissues (e.g., blood vessels and muscles) with varying sizes and shapes.

Our main contributions are threefold. First, we propose a novel DONet containing multiple Dual-OctConvs connected in a dense manner for multifrequency feature reuse and demonstrate their ability to capture multiscale contextual information. Second, we employ the Dual-OctConv to deal with complex-valued inputs in a multifrequency representation space and encourage information exchange and fusion across various frequency domains. The Dual-OctConv is a generalization of the standard complex convolution and endows our model with several appealing characteristics (e.g., a larger receptive field, higher flexibility, and more computational efficiency). Third, our model shows significant performance improvements against state-of-the-art algorithms on a clinical knee dataset and the fastMRI dataset.

Compared with our preliminary conference version [34], this article provides the following extensions.

- 1) We generalize the architecture of the dual-octave (corresponding to DONet in this article) from [34] to a more powerful version (i.e., DONet). This is done by employing dense connections for better feature reuse. Our DONet is demonstrated to achieve significantly better performance than dual-octave [34] (see Section III).
- 2) We further apply the proposed DONet to the currently largest raw MRI dataset, fastMRI, to demonstrate the advantages of our method over other deep learning models (see Section IV).
- 3) Finally, we also provide a more inclusive and insightful overview of recent work on fast parallel MRI reconstruction (see Section II).

II. RELATED WORK

A. Deep Learning in MR Image Reconstruction

Ever since the pioneering works introducing CNNs for computer vision tasks, such as image classification and face
recognition, researchers have made substantial efforts to improve medical and clinical practice using deep learning techniques [35], [36]. Wang et al. [31] proposed the first deep learning-based MR image reconstruction framework, which learns the mapping between fully sampled single-coil MR images and their counterpart data reconstructed from a zero-filled undersampled k-space. A large number of networks have since been developed for MR image reconstruction, especially nonparallel reconstruction [37]. For example, Yang et al. [38] proposed a model-based unrolling method that formulates the algorithm within a deep neural network and trained the network with a small amount of data. As an end-to-end method, Han et al. [39] employed U-Net to model a domain adaptation structure that removes aliasing artifacts from corrupted images. Zhu et al. [40] proposed the automated transform by manifold approximation (AUTOMAP) framework, which they implemented using a deep neural network to learn a mapping between the sensor and image domain. Han et al. [39] proposed a deep learning approach with domain adaptation, which uses a large amount of X-ray CT or synthesized radial MR data and is fine-tuned with a small amount of real radial MR data. Lee et al. [41] trained amplitude and phase, respectively, using a residual network to reduce the computational cost of the reconstruction algorithm. With the pioneering works on generative adversarial networks (GANs) for natural image synthesis, GAN-based methods have also made great progress in MR image reconstruction [42]–[44]. For example, Quan et al. [42] designed an adversarial model for CS-MR image reconstruction, accurately interpolating and restoring data in an undersampled k-space data, and with a loss of data consistency. Moreover, image refinement on specific network structures with different training objectives, such as the adversarial loss [43] or perceptual loss [42], has gained more attention.

In parallel imaging, one representative model is the variational network (VN-Net) [25], which combines the mathematical structure of the variational model with deep learning for fast multi-coil MR image reconstruction. Another model-based deep framework [26] was designed with a split Bregman iterative algorithm to achieve accurate reconstruction from multi-coil undersampled k-space data. To obtain high-fidelity reconstructions, GrappaNet [30] was proposed to combine traditional parallel imaging methods with deep neural networks. Recently, complex-valued representations have demonstrated superiority in processing complex-valued inputs [32]. For example, Wang et al. [2] applied complex convolutions to jointly process real and imaginary values for comprehensive feature representations. In contrast, our approach represents complex-valued input features in a multifrequency space. The Dual-OctConv, proposed for processing such multifrequency data, can capture richer contextual knowledge, leading to significant improvement in performance.

B. Multiscale Representation Learning

Multiscale information has proven effective in various computer vision tasks (e.g., image classification, object detection, and semantic segmentation). Especially in the era of deep learning, multiscale representation has been successfully applied in various fields due to its strong robustness and generalization ability. Several strategies have been proposed for this, yielding significant performance improvement in a number of tasks. For example, Ke et al. [45] proposed a multigrid network to propagate and integrate information across multiple scales for image classification. Multiscale information has also been proven effective in restoring image details for image enhancement [46]–[48]. Nah et al. [46] proposed a multiscale CNN together with a multiscale loss function for dynamic scene deblurring. In addition, various well-known techniques (e.g., FPN [49] and PSP [50]) have been proposed for learning multiscale representations in object detection and segmentation tasks. For example, Lin et al. [49] proposed a top-down feature pyramid architecture with lateral connections at all scales for object detection. Recently, the octave convolution [33] was proposed to learn multiscale features based on the spatial frequency model [51], [52], greatly improving performance in natural image and video recognition.

In this work, we demonstrate the appealing properties of the octave convolution for accelerated parallel MR image reconstruction, which helps to capture multiscale information from features of multiple spatial frequencies. Based on this, we propose a novel Dual-OctConv for accelerated parallel MR image reconstruction, which enables our model to capture details of vasculatures and tissues with varying sizes and shapes, yielding high-fidelity reconstructions.

III. METHODOLOGY

A. Problem Formulation

MR scanners acquire k-space data through the receiver coils and then utilize an inverse multidimensional Fourier transform to obtain the final MR images. In parallel imaging, multiple receiver coils are used to simultaneously acquire k-space data from the target under scanning.

Let \( A = MF \in \mathbb{C}^{M \times N} \) denote the undersampled Fourier encoding matrix, where \( F \) is the multidimensional Fourier transform and \( M \) is an undersampled mask operator. In parallel imaging, the same mask is used for all coils. The undersampled k-space data from each coil can be expressed as

\[
y_i = A(S_i x)
\]

where \( i = 1, 2, \ldots, c \), with \( c \) denoting the number of coils, \( x \in \mathbb{C}^{N \times 1} \) is the ground-truth MR image, \( y_i \in \mathbb{C}^{M \times 1}(M \ll N) \) is the undersampled k-space data for the \( i \)th coil, and \( S_i \) is a complex-valued diagonal matrix encoding the sensitivity map of the \( i \)th coil. The coil sensitivities, which are measured by each coil, modulate the k-space data. The coil configuration and interactions with the anatomical structures under scanning can affect the coil sensitivities, so \( S_i \) changes across different scans. In addition, the obtained image will contain aliasing artifacts if the inverse Fourier transform (IFT) is directly applied to the undersampled k-space data. For single-coil MR imaging, \( i = 1 \).

We can reconstruct \( \hat{x} \) with prior knowledge of its properties, which can be formulated as the following...
optimization problem:

$$\hat{x} = \arg \min_{x} \sum_{i=1}^{c} \| y_i - A(S, x) \|^2_2 + \lambda \Psi(x)$$  \hspace{1cm} (2)$$

where $\Psi$ is a regularization function and $\lambda$ controls the tradeoff between the two terms.

The problem presented in (2) can be effectively resolved using CNNs, which avoids time-consuming numerical optimization and the need for a coil sensitivity map. During training, we update the network weights by minimizing an $\ell_1$ loss function

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{N} \sum_{n=1}^{N} \| x'(n) - f_\theta(y'(n)) \|_1$$  \hspace{1cm} (3)$$

where $y'(n)$ is the $n$th multichannel image obtained from the zero-filled $k$-space data, $x'(n)$ is the $n$th multichannel ground truth image, $N$ is the total number of training samples, and $f_\theta(\cdot)$ is an end-to-end mapping function parameterized by $\theta$, which contains a large number of adjustable network weights. Training with (3) can reconstruct the expected MR images, but the original information of the data acquired in the $k$-space cannot be well preserved. If we incorporate the undersampled $k$-space data into the data fidelity at the training stage, the network can yield improved reconstruction results. For this purpose, we add the data fidelity units in our network, as in [2]. After the network is trained, we obtain a set of optimal parameters $\hat{\theta}$ for the reconstruction of a multichannel image and predict this image via $\hat{x}' = f_{\hat{\theta}}(y')$. Finally, we use an adaptive coil combination method [2] to obtain the expected MR image from $\hat{x}'$.

\[ B. \text{ Framework Overview} \]

Fig. 2 illustrates the architecture of our DONet, which simultaneously considers both $k$-space data fidelity and image space proximity to achieve high-fidelity reconstruction. Specifically, given an undersampled multicoil $k$-space measurement as input, we first convert it to aliased multichannel images via an IFT. The transformed image is then fed into a cascade of $T$ Dual-OctConv blocks, with each block followed by a data fidelity unit [2]. The data fidelity unit helps to preserve the original $k$-space information during training. Formally, the $t$th data fidelity unit can be written as

$$X' = F^{-1}((1 - M)F(\hat{X}') + My)$$  \hspace{1cm} (4)$$

where $\hat{X}'$ denotes the reconstructed output image of the $t$th Dual-OctConv block, while $X'$ represents the output after applying the $t$th data fidelity unit, which will be fed into the $(t+1)$th Dual-OctConv block as input. $F^{-1}$ denotes the IFT.

Next, we will describe the details of our Dual-OctConv block in Section III-C and the Dual-OctConv layer in Section III-D.

\[ C. \text{ Dual-Octave Convolutional Block} \]

We present the detailed architecture of our Dual-OctConv block, which is employed to learn comprehensive, multiscale feature representations for reconstruction. In the following paragraphs, the superscript “$\ell$” is omitted for conciseness, unless necessary.

As shown in Fig. 3, each Dual-OctConv block consists of $K$ Dual-OctConv layers $\{D_k\}_{k=1}^{K}$. These layers accept $X$ as input and produce the reconstruction $\hat{X}$ in a sequential manner. More specifically, to obtain rich multiscale context information, we first represent the multichannel input with complex filters and then decompose it into low and high spatial frequency parts. Let $X \in \mathbb{C}^{c \times h \times w}$ be the input complex feature maps with $c$, $h$, and $w$ denoting the number of channels, height, and width, respectively. $X_{r}(0)$ and $X_{h}(0)$ are the real and imaginary parts of the complex data, respectively. As illustrated in Fig. 3, we initially split the complex input feature maps into low- and high-frequency groups: $X_{r}(0) = \{X_{r}^{L}(0), X_{r}^{H}(0)\}$ and $X_{h}(0) = \{X_{h}^{L}(0), X_{h}^{H}(0)\}$, where $X_{r}^{H} \in \mathbb{C}^{1-\alpha c \times h \times w}$ captures the high-frequency fine details of the data and $X_{L} \in \mathbb{C}^{\alpha c \times h \times w}$ determines the low-frequency image contrast. Here, $\alpha \in [0, 1]$ controls the ratio of channels that are allocated to low-frequency and high-frequency feature maps. Note that the Dual-OctConv becomes the standard complex convolution [32] when $\alpha = 0$. Then, for the $k$th Dual-OctConv layer $D_k$, we have

$$X_{r}^{L}(k), X_{r}^{H}(k), X_{h}^{L}(k), X_{h}^{H}(k) = D_k(X_{r}^{L}(k-1), X_{r}^{H}(k-1), X_{h}^{L}(k-1), X_{h}^{H}(k-1))$$  \hspace{1cm} (5)$$

where $X_{r}^{L}(k), X_{r}^{H}(k-1), X_{h}^{L}(k-1)$, and $X_{h}^{H}(k-1)$ are the outputs, and $X_{r}^{L}(k), X_{r}^{H}(k), X_{h}^{L}(k), X_{h}^{H}(k)$ are the inputs of $D_k$. Then, for each block, we obtain the final feature representations at the $K$th Dual-OctConv layer, which are combined to obtain the reconstruction

$$\hat{X} = u(c(X_{r}^{L}(K), X_{h}^{L}(K)), 2) + c(X_{r}^{H}(K), X_{h}^{H}(K))$$  \hspace{1cm} (6)$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where \( u(\cdot, z) \) denotes the upsampling operation with a factor of \( z \). Here, we use the nearest neighbor interpolation. \( c(\cdot, \cdot) \) denotes the concatenation operation. The output \( \mathbf{X} \) is then fed into the data fidelity layer [see (4)] and the subsequent Dual-OctConv block for cascade reconstruction.

In addition, we emphasize that the sequential pipeline mentioned above conducts individual feature learning by layer, neglecting the feature fusion among different layers, which often produces more comprehensive feature representations [53], [54]. To this end, we improve the sequential Dual-OctConv block by introducing dense connections in an interleaved manner. Formally, for the output of the \( k \)-th \((k \geq 1)\) Dual-OctConv layer, its features are modulated as follows:

\[
\begin{align*}
\mathbf{X}^L_{r(k)} &= C_k([\mathbf{X}^L_{r(k-1)}, \mathbf{X}^L_{r(k-1)}, \ldots, \mathbf{X}^L_{r(0)}]), \\
\mathbf{X}^L_{i(k)} &= C_k([\mathbf{X}^L_{i(k)}, \mathbf{X}^L_{i(k-1)}, \ldots, \mathbf{X}^L_{i(0)}]), \\
\mathbf{X}^H_{r(k)} &= C_k([\mathbf{X}^H_{r(k)}, \mathbf{X}^H_{r(k-1)}, \ldots, \mathbf{X}^H_{r(0)}]), \\
\mathbf{X}^H_{i(k)} &= C_k([\mathbf{X}^H_{i(k)}, \mathbf{X}^H_{i(k-1)}, \ldots, \mathbf{X}^H_{i(0)}]).
\end{align*}
\]

(7)

Here, the feature maps at the \( k \)-th layer have first concatenated the features at all previous layers and then processed with a function \( C_k \) to learn collective knowledge from all these feature maps. Following Huang et al. [54], we implement \( C_k \) as BN-ReLU-Conv(1 \times 1)-BN-ReLU-Conv(3 \times 3). Such dense connections enable our model to reuse features from previous layers with high computational efficiency since there is no need to relearn redundant feature maps. They also help to alleviate the gradient vanishing problem when the number of layers in our network increases. In our experiments, we demonstrate that the reconstruction performance significantly improves after introducing these dense connections.

**D. Dual-Octave Convolutional Layer**

Now, we present the proposed dual-octave convolutional layer, which fuses the real and imaginary parts of a complex valued input. For the \( k \)-th layer \( D_k \), we denote \( \mathbf{X}_{r(k-1)} = \mathbf{X}_{r(k-1)} + i\mathbf{X}_{i(k-1)} \) as its input. We convolve \( \mathbf{X}_{r(k-1)} \) with a complex filter matrix \( \mathbf{K}_{r(k)} = \mathbf{K}_{r(k)} + i\mathbf{K}_{i(k)} \) as follows:

\[
\begin{align*}
\mathbf{K}_{r(k)} &\text{ and } \mathbf{K}_{i(k)} \text{ represent real and imaginary kernels, respectively. Note that all the kernels and feature maps are expressed by real matrices since the complex arithmetics are simulated by real-valued entities.}
\end{align*}
\]

As shown in Fig. 4, the complex filter matrix is further expressed as \( \mathbf{K}_{r(k)} = \text{[} \mathbf{K}^H_{r(k)}, \mathbf{K}^L_{r(k)}, \text{]} \) and \( \mathbf{K}_{i(k)} = \text{[} \mathbf{K}^H_{i(k)}, \mathbf{K}^L_{i(k)}, \text{]} \) convolve with \( \mathbf{X}_{r(k-1)} \), \( \mathbf{X}_{i(k-1)} \), \( \mathbf{X}^H_{r(k-1)} \), and \( \mathbf{X}^H_{i(k-1)} \). Then, \( D_k \) can be formulated as:

\[
\begin{align*}
\mathbf{X}_{r(k)} &= f(\mathbf{X}_{r(k-1)}; \mathbf{K}^L_{r(k)}, \mathbf{K}^L_{r(k)-1}) + u(\mathbf{f}(\mathbf{X}_{r(k-1)}; \mathbf{K}^H_{r(k)}, 2)) + f(\mathbf{X}_{i(k-1)}; \mathbf{K}^L_{i(k)}, \mathbf{K}^L_{i(k)-1}); \\
\mathbf{X}_{i(k)} &= f(\mathbf{X}_{i(k-1)}; \mathbf{K}^L_{i(k)}, \mathbf{K}^L_{i(k)-1}) + u(\mathbf{f}(\mathbf{X}_{i(k-1)}; \mathbf{K}^H_{i(k)}, 2)) + f(\mathbf{X}_{r(k-1)}; \mathbf{K}^L_{r(k)}, \mathbf{K}^L_{r(k)-1}); \\
\mathbf{X}^H_{r(k)} &= f(\mathbf{X}^H_{r(k-1)}; \mathbf{K}^H_{r(k)}, \mathbf{K}^H_{r(k)-1}) + f(p(\mathbf{X}^H_{r(k-1)}); \mathbf{K}^H_{r(k-1)}); \\
\mathbf{X}^H_{i(k)} &= f(\mathbf{X}^H_{i(k-1)}; \mathbf{K}^H_{i(k)}, \mathbf{K}^H_{i(k)-1}) + f(p(\mathbf{X}^H_{i(k-1)}); \mathbf{K}^H_{i(k-1)}),
\end{align*}
\]
where $f(X; K)$ indicates the convolution with kernel $K$, $u(X, k)$ denotes the upsampling operation with a factor of $k$ via nearest interpolation, and $p(X, z)$ represents an average pooling layer with kernel size $z \times z$. The real and imaginary parts are fused with the operations $\{L \rightarrow L, H \rightarrow H\}$ and $\{H \rightarrow L, L \rightarrow H\}$, which corresponds to the information updating and exchanging between high- and low-frequency feature maps. Therefore, our Dual-OctConv is able to enlarge the receptive fields of the low-frequency feature maps both in the real and imaginary parts. To put this into perspective, after convolving the low-frequency feature maps of the real and imaginary parts ($X_{r(k-1)}^L$, $X_{r(k-1)}^R$) with $z \times z$ complex convolutional kernels, the receptive fields of both achieve a $2 \times$ enlargement compared to the vanilla convolution. Thus, our Dual-OctConv has a strong ability to capture rich context information at different scales.

IV. EXPERIMENTS

A. Datasets

We introduce a clinical knee [25] dataset and the fastMRI [55] dataset under different undersampling patterns and acceleration factors to evaluate our method. The clinical multicoil fully sampled MR knee dataset is acquired using a clinical 3T Siemens Magnetom Skyra scanner with a sequence called “Coronal Spin Density Weighted Without Fat Suppression.” The imaging protocol is detailed as follows: the 15-channel knee coil, the matrix size $320 \times 320 \times 20$, TR = 2750 ms, TE = 27 ms, and in-plane resolution = $0.49 \times 0.44$ mm$^2$. There are 20 subjects in total with the following information: five female/15 male, age 15–76, and BMI 20.46–32.94. We randomly select fourteen patients for training, three for validation, and three for testing. For the fastMRI dataset, we randomly select 400 patients from the multicoil knee dataset using a proton density (PD) weighted sequence for training, 40 for validation, and 60 for testing. In our experiments, we exclude the first few slices of each volume since the frontal slices are much noisier than the other slices, making the distribution of frontal slices different from the rest. More details of the fastMRI dataset can be obtained from [55].

The predefined undersampling masks are used to obtain the undersampled measurements. In our experiments, we adopt four different $k$-space undersampling patterns, including 1-D uniform, 1-D Cartesian, 2-D random, and 2-D radial. Examples of the undersampling patterns are illustrated in Figs. 5 and 6. For 1-D uniform, 1-D Cartesian, and 2-D random masks, the acceleration rate is set to $3 \times$ and $5 \times$. For the 2-D radial mask, $4 \times$ and $6 \times$ accelerations are adopted.

B. Implementation Details

We implement our model using Tensorflow 1.14 and perform experiments using an NVIDIA 1080Ti GPU with 11 GB
of memory. Following Wang et al. [2], we initialize the magnitude and phase of the complex parameters using Rayleigh and uniform distributions, respectively. The network is trained using the Adam optimizer [2] with an initial learning rate of 0.001 and a weight decay of 0.95. The batch size is set to four, and the convolutional kernel size is set to $3 \times 3$. Each complex convolutional layer has 64 feature maps, except for the last layer, which is determined by the concatenated real and imaginary channels of the data. The spatial frequency ratio $\alpha$ is set to 0.125 by default.

To demonstrate their effectiveness, we compare our DONet and DONet$^\dagger$ (without dense connections) with several state-of-the-art parallel MR imaging approaches, including traditional methods (SPIRiT [20] and L1-SPIRiT [56]) and CNN-based methods (VN-Net [25] and ComplexMRI [2]). All these methods are trained on the same dataset with their default settings. For CNN-based methods, we retrain them according to the specifications with TensorFlow, using their default parameter settings.

C. Quantitative Evaluation

We use peak SNR (PSNR) and structural similarity index measure (SSIM) [2] for quantitative evaluation. Table I reports the average PSNR and SSIM results with respect to different undersampling patterns and acceleration factors on the clinical dataset. As can be seen, DONet$^\dagger$ obtains consistent performance improvements against all baseline methods, demonstrating the superiority of our Dual-OctConv. Moreover, by incorporating dense connections, our full model (i.e., DONet) further improves the performance significantly across various settings.

In addition, we observe that the undersampling patterns greatly affect the quality of reconstruction. For instance, the 2-D sampling masks generally outperform the 1-D masks. Another important observation is that the reconstruction becomes more difficult when the acceleration rate increases. In particular, our model significantly outperforms previous methods under extremely challenging settings (e.g., 2-D masks with $5 \times$ and $6 \times$ accelerations). This can be attributed to the powerful capability of our DONet in aggregating rich contextual information of real and imaginary data. Moreover, the overall results in Table I show the strong robustness of our model under various undersampling patterns and acceleration rates. For example, our method restores more information and with minimum artifacts, as demonstrated in Figs. 5 and 6 with error maps.

In Table II, we provide reconstruction evaluations of our model on fastMRI, which is currently the largest MR image dataset. As we can see, DONet obtains the best reconstruction results among different undersampling patterns and acceleration rates. In particular, under 1-D uniform undersampling patterns with $5 \times$ acceleration, our DONet improves the reconstruction PSNR from 29.992 to 32.780 dB and SSIM from
TABLE I
QUANTITATIVE COMPARISON OF STATE-OF-THE-ART METHODS UNDER DIFFERENT UNDERSAMPLING PATTERNS ON THE CLINICAL KNEE DATASET. BEST AND SECOND BEST RESULTS ARE MARKED IN RED AND BLUE, RESPECTIVELY

| undersampling pattern     | 1D Uniform 3x | 1D Uniform 5x | 1D Cartesian 3x | 1D Cartesian 5x | 2D Random 3x | 2D Random 5x | 2D Radial 4x | 2D Radial 6x |
|---------------------------|---------------|---------------|------------------|------------------|---------------|---------------|---------------|---------------|
|                           | PSNR          | SSIM          | PSNR             | SSIM             | PSNR          | SSIM          | PSNR          | SSIM          |
| Zero-filling              | 24.406.676 | 23.579.655 | 25.922.726 | 24.550.685 | 30.540.827 | 27.078.750 | 31.026.826 | 28.107.766 |
| SPIRIT                    | 29.385.700 | 28.300.676 | 32.310.801 | 31.222.782 | 32.179.786 | 32.258.812 | 30.308.720 | 29.061.702 |
| L1-SPIRIT                 | 29.815.847 | 27.353.788 | 33.346.887 | 30.912.837 | 38.597.937 | 34.071.887 | 37.004.919 | 34.149.881 |
| VN-Net                    | 35.436.907 | 32.730.858 | 36.364.912 | 33.236.866 | 38.409.956 | 35.734.923 | 37.956.930 | 34.609.907 |
| ComplexMRI                | 34.989.909 | 32.803.873 | 35.957.916 | 34.126.876 | 39.563.946 | 37.315.908 | 38.098.933 | 35.768.904 |
| DONet†                    | 36.243.919 | 34.128.885 | 37.029.923 | 34.944.884 | 39.964.948 | 38.279.930 | 38.607.935 | 36.584.908 |
| DONet                     | 36.338.918 | 34.526.892 | 37.074.931 | 35.400.902 | 40.652.953 | 38.428.930 | 38.387.935 | 35.281.909 |

0.829 to 0.856 compared to the best previous state-of-the-art fusion method, ComplexMRI. Similarly, we observe that our DONet yields outstanding results on both the 1-D uniform and 1-D Cartesian undersampling patterns with 3× and 5× accelerations. More importantly, even without the dense connection, our DONet† achieves compelling reconstruction results compared to the state-of-the-art methods. Under the 2-D radial pattern with 3× acceleration, DONet† achieves SSIM = 0.930. All the results demonstrate that our DONet preserves powerful and rich contextual information for real and imaginary data.

D. Qualitative Evaluation

For the qualitative analysis, we first show the reconstructed images and corresponding error maps for 1-D uniform and 1-D Cartesian sampling with a 3× acceleration rate in Fig. 5. In general, 1-D Cartesian masks provide better results in comparison to 1-D uniform masks. Our method provides the best-quality reconstructed images and significantly reduces prediction errors. In contrast, the baseline methods yield large prediction errors and show unsatisfactory performance.

We next examine the results for 2-D radial masks with a 4× acceleration rate and 2-D random masks with a 5× acceleration rate. As shown in Fig. 6, CNN-based methods significantly improve the results with fewer errors and clearer structures in comparison with SPIRIT and L1-SPIRIT. In particular, our DONet produces higher-quality images with clear details and minimum artifacts. The superior performance is due to the fact that our method can effectively aggregate the information of various spatial frequencies present in the real and imaginary parts of an MR image.

E. Ablation Studies

The crucial part of our DONet is the Dual-OctConv. It is, thus, important to prove the effectiveness of the proposed convolution. First, we study the effects of Dual-OctConv. For comparison, we build a baseline convolution by setting \( \alpha = 0 \),
which turns our method into a standard complex convolution. We conduct experiments on the test set of in vivo dataset with 60 complex-valued images under the uniform undersampling mask with a $3 \times$ acceleration rate. As illustrated in Fig. 7, our Dual-OctConv significantly outperforms the baseline model, especially in terms of SSIM. This reveals the superiority of the proposed method in improving the reconstruction.

Second, we investigate the influence of the spatial frequency ratio $\alpha$ for reconstruction. The ratio determines the receptive fields in both the real and imaginary parts and also influences the fusion of these parts at multiple spatial frequencies. As shown in Fig. 8, our model achieves the best PSNR and SSIM scores at $\alpha = 0.125$, which means that 12.5% of the channels in the real and imaginary parts are reduced to a low spatial frequency. When $\alpha$ becomes larger, the performance quickly degrades due to severe information loss induced by overlarge ratios.

The number of network parameters increases as the number of blocks (bn) increases. Therefore, it is necessary to choose an appropriate number of blocks to ensure that the network structure reaches the highest reconstruction accuracy without inducing higher computational and memory requirements. Herein, we carry out various experiments using different numbers of blocks. The results are presented in Fig. 9. As can be seen from the curves, our model can successfully reconstruct the MR images at bn = 4, and the reconstruction accuracy reaches the highest at bn = 10.

Finally, we study the floating-point operations per seconds (FLOPs) of DONet with respect to different $\alpha$’s in Fig. 10. The number below each point is the value of $\alpha$. As can be observed, a small $\alpha$ leads to improved performance with higher FLOPs. Moreover, compared with the baseline model (e.g., $\alpha = 0$), our model consistently shows better performance with much lower FLOPs.

V. CONCLUSION

In this work, we focus on spatial frequency feature expression in complex-valued data for parallel MR image reconstruction. For this purpose, we propose a DONet with a series of novel Dual-OctConv operations to deal with the real and imaginary components of the data at multiple spatial frequencies. By convolving the feature maps of both the real and imaginary components under different spatial resolutions, the proposed Dual-OctConv facilitates our DONet to learn more comprehensive feature representations, yielding higher-quality reconstructed images with significantly reduced artifacts. We conduct extensive experiments on the in vivo knee and fastMRI datasets under different settings of undersampling patterns and acceleration rates. The results demonstrate the advantages of our model against state-of-the-art methods in accelerated MR image reconstruction.

REFERENCES

[1] L. Sun, Z. Fan, Y. Huang, X. Ding, and J. Paisley, “Compressed sensing MRI using a recursive dilated network,” in Proc. 32nd AAAI Conf. Artif. Intell., 2018, pp. 1444–1451.
[2] S. Wang et al., “DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution,” Magn. Reson. Imag., vol. 68, pp. 136–147, May 2020.
[3] H. K. Aggarwal, M. P. Mani, and M. Jacob, “MoDL: Model-based deep learning architecture for inverse problems,” IEEE Trans. Med. Imag., vol. 38, no. 2, pp. 394–405, Feb. 2019.
[4] C.-M. Feng, K. Wang, S. Lu, Y. Xu, and X. Li, “Brain MRI super-resolution using coupled-projection residual network,” Neurocomputing, vol. 456, pp. 190–199, Oct. 2021.
[5] S. G. Lingala and M. Jacob, “Blind compressive sensing dynamic MRI,” IEEE Trans. Med. Imag., vol. 32, no. 6, pp. 1132–1145, Jun. 2013.
[6] Y. Wang and L. Ying, “Undersampled dynamic magnetic resonance imaging using kernel principal component analysis,” in Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Aug. 2014, pp. 1533–1536.
[7] E. Hot and P. Sekulic, “Compressed sensing MRI using masked DCT and DFT measurements,” in Proc. 4th Medit. Conf. Artif. Intell. Imag. Proc. 32nd AAAI Conf. Artif. Intell., 2018, pp. 1444–1451.
[8] A. C. Gilbert, P. Indyk, M. Iwen, and L. Schmidt, “Recent developments in the sparse Fourier transform: A compressed Fourier transform for big data,” IEEE Signal Process. Mag., vol. 31, no. 5, pp. 91–100, Sep. 2014.
[9] Z. Lai et al., “Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform,” Med. Image Anal., vol. 27, pp. 93–104, Jan. 2016.
Chun-Mei Feng received the M.S. degree from Qufu Normal University, Jining, China, in 2018. She is currently pursuing the Ph.D. degree with the School of Information Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China. Her research interests include medical imaging and bioinformatics.

Zhanyuan Yang received the M.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2020. His research interests are computer vision based on deep learning, including image restoration and image enhancement.

Huazhu Fu (Senior Member, IEEE) received the Ph.D. degree from Tianjin University, Tianjin, China, in 2013. He was a Research Fellow with Nanyang Technological University, Singapore, for two years. From 2015 to 2018, he was a Research Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. He is currently a Senior Scientist with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates. His research interests include computer vision, machine learning, and medical image analysis. Dr. Fu also serves as the Co-Chair of the Ophthalmic Medical Image Analysis (OMIA) International Workshop and a Co-Organizer of the Ocular Image Series Challenge (i-Challenge). He also serves as an Associate Editor for IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, and IEEE ACCESS.

Yong Xu (Senior Member, IEEE) received the B.S. and M.S. degrees from the Air Force Institute of Meteorology, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, Nanjing, in 2005. From May 2005 to April 2007, he was a Post-Doctoral Research Fellow with the Harbin Institute of Technology (Shenzhen) (HIT Shenzhen), Shenzhen, China, where he is currently a Professor. His current interests include pattern recognition, machine learning, and bioinformatics.

Jian Yang (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2002. In 2003, he was a Post-Doctoral Researcher with the University of Zaragoza, Zaragoza, Spain. From 2004 to 2006, he was a Post-Doctoral Fellow with the Biometrics Centre, The Hong Kong Polytechnic University, Hong Kong. From 2006 to 2007, he was a Post-Doctoral Fellow with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA. He is currently the Chang-Jiang Professor with the School of Computer Science and Technology, NUST. He is the author of more than 200 scientific articles in pattern recognition and computer vision. His articles have been cited more than 6000 times in the Web of Science and 18 000 times in the Scholar Google. His research interests include pattern recognition, computer vision, and machine learning. Dr. Yang is also a fellow of the International Association for Pattern Recognition (IAPR). He has been an Associate Editor of Pattern Recognition, Pattern Recognition Letters, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and Neurocomputing.

Ling Shao (Fellow, IEEE) is currently the CEO and the Chief Scientist with the Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates. His research interests include computer vision, machine learning, and medical imaging. Dr. Shao is also a fellow of the International Association for Pattern Recognition (IAPR), the Institution of Engineering and Technology (IET), and the British Computer Society (BCS). He is also an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and several other journals.