Attention hybrid variational net for accelerated MRI reconstruction

Cite as: APL Mach. Learn. 1, 046116 (2023); doi: 10.1063/5.0165485
Submitted: 28 June 2023 • Accepted: 15 November 2023 •
Published Online: 8 December 2023

Guoyao Shen,[1] Boran Hao,[2] Mengyu Li,[1] Chad W. Farris,[3] Ioannis Ch. Paschalidis,[2] Stephan W. Anderson,[3] and Xin Zhang[1,a)]

AFFILIATIONS

[1] Department of Mechanical Engineering and the Photonics Center, Boston University, Boston, Massachusetts 02215, USA
[2] Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215, USA
[3] Boston Medical Center and Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts 02118, USA

[a)] **Author to whom correspondence should be addressed:** xinz@bu.edu

ABSTRACT

The application of compressed sensing (CS)-enabled data reconstruction for accelerating magnetic resonance imaging (MRI) remains a challenging problem. This is due to the fact that the information lost in k-space from the acceleration mask makes it difficult to reconstruct an image similar to the quality of a fully sampled image. Multiple deep learning-based structures have been proposed for MRI reconstruction using CS, in both the k-space and image domains, and using unrolled optimization methods. However, the drawback of these structures is that they are not fully utilizing the information from both domains (k-space and image). Herein, we propose a deep learning-based attention hybrid variational network that performs learning in both the k-space and image domains. We evaluate our method on a well-known open-source MRI dataset (652 brain cases and 1172 knee cases) and a clinical MRI dataset of 243 patients diagnosed with strokes from our institution to demonstrate the performance of our network. Our model achieves an overall peak signal-to-noise ratio/structural similarity of 40.92 ± 0.29/0.9577 ± 0.0025 (fourfold) and 37.03 ± 0.25/0.9365 ± 0.0029 (eightfold) for the brain dataset, 31.09 ± 0.25/0.6901 ± 0.0094 (fourfold) and 29.49 ± 0.22/0.6197 ± 0.0106 (eightfold) for the knee dataset, and 36.32 ± 0.16/0.9199 ± 0.0029 (20-fold) and 33.70 ± 0.15/0.8882 ± 0.0035 (30-fold) for the stroke dataset. In addition to quantitative evaluation, we undertook a blinded comparison of image quality across networks performed by a subspecialty trained radiologist. Overall, we demonstrate that our network achieves a superior performance among others under multiple reconstruction tasks.

I. INTRODUCTION

Magnetic resonance imaging (MRI) is a powerful tool in clinical medicine and research settings, from diagnosing knee injuries to studying brain disease. However, one major challenge in MRI is the long data acquisition time, ultimately limiting patient access to this diagnostic tool. Furthermore, the long image acquisition times also lead to patient discomfort, worsening potential claustrophobia, an increased chance of patient motion, and degraded images that reduce diagnostic utility. In order to reduce the scan time and improve the efficiency, compressed sensing (CS) techniques[1,2] that enable less measurement data in k-space in order to reconstruct the final image have been developed and successfully applied to MRI. However, these under-sampling strategies come at a cost in image quality with resultant image blurring and aliasing artifacts that can significantly influence the diagnostic yield of MRI.

Recently, thanks to achievements in deep learning, especially the successful application of deep convolutional neural networks (DCNNs) in multiple imaging tasks,[3] a new paradigm has been realized, in which image reconstruction can be accomplished by exploiting the latent representation within the neural network.[4–7] Importantly, these deep learning-based strategies also provide an additional benefit whereby their network structure requires no or minimal modification before they may be applied to different tasks. For a typical CS reconstruction task in MRI, there are two major domains of relevance: the k-space domain and the image domain. Data under-sampling typically occurs in k-space such that a mask is applied and k-space is incompletely sampled. If Fourier

transformation is then applied to these under-sampled k-space data with all the masked portions zero-filled, this is referred to as zero-filled MRI (ZF-MRI). The overall quality of ZF-MRI relies on the degree of under-sampling and the pattern of the k-space mask.[1,8] There are ongoing efforts focused on reconstructing a high-quality image using the under-sampled k-space, or ZF-MRI.

To this end, multiple deep learning structures have been proposed to leverage the CS reconstruction task and improve the ultimate image quality. U-Net is a well-known structure that was proposed to handle image segmentation tasks[9] but was later reported to be feasible for MRI reconstruction.[10,11] The U-Net structure focused on reconstruction in the image domain, employing a ZF-MRI as input and reconstructing an unaliased image. It has also been shown that employing a data consistency (DC) layer[12] serves to improve the performance of a deep convolutional neural network (DCNN) in the image domain due to the reuse of the visible portions of k-space. Furthermore, with an increase in interest in the attention mechanism in DCNNs, spatial- and channel-wise attention approaches[13–17] have also been reported to provide additional benefits in reconstruction performance.

In contrast to methods focused on the image domain, automated transform by manifold approximation (Automap)[18–20] is a method focused on learning the domain transformation in order to perform reconstruction directly from the k-space domain. The fully connected layers encode the transformation of the complex data into image space, followed by a convolution-deconvolution structure that provides the reconstruction of the final image. Even though this method provides a prototype of learning from the combination of both the k-space and image space domains, one major disadvantage of this method is its sensitivity to image size. In the Automap technique, the dimensions of the fully connected layers are closely related to the image shape, making it difficult to train for a larger image.

Unrolled optimization-based structures, on the other hand, are a type of network that divides the overall reconstruction task into multiple steps, where each step contains one or multiple sub-networks.[21] Variational net (VarNet)[22,23] is one such unrolled optimization-based structure that utilizes not only the under-sampled k-space data but also the mask and sensitivity maps during intermediate states.

W-Net (Double U-Net) is a hybrid network that performs reconstruction in both the k-space and image domains.[24] The network consists of two U-Net structures and an inverse Fourier transformation layer connecting them. It has been shown that this dual-domain learning structure provides a better reconstruction performance compared to a single-domain learning structure. Yet the W-Net lacks in k-space learning efficiency as the k-space domain U-Net holds a small weight in the combined loss function, putting more weight on image-domain learning.

In this work, we seek to leverage and further develop a CS reconstruction method. More specifically, we build an attention hybrid variational network (AttHybrid-VarNet) that benefits from the superior k-space reconstruction ability and an image-domain refinement network to further improve the image quality. Furthermore, spatial- and channel-wise attention also enables the convolutional module to further fine tune the weights for different channels and regions in the feature maps according to the attention scores. We compare our architecture with multiple CS reconstruction networks over open-source and clinical imaging datasets. We also provide a blinded radiologist evaluation of the image quality of our methods compared to other typical reconstruction networks. Ultimately, we demonstrate that our network achieves a superior performance among others under multiple reconstruction setups.

II. MATERIALS AND METHODS

A. End-to-end variational network

Variational networks have shown a superior performance for MRI reconstruction tasks. Consider an MR image acquiring measurement,

$$\boldsymbol{k} = \mathcal{F}(\boldsymbol{x}) + e, \tag{1}$$

where $\boldsymbol{x}$ is the underlying image, $\boldsymbol{k}$ is the k-space measurement, $\mathcal{F}$ is the Fourier transform operator, and e stands for the measurement noise. In an accelerated MRI acquisition case, $\tilde{\boldsymbol{k}} = M\boldsymbol{k}$, where M is the binary under-sampling mask applied to the k-space, and $\tilde{\boldsymbol{k}}$ denotes the under-sampled k-space measurement.

An estimation of the underlying image can be solved using the following optimization,

$$\hat{\boldsymbol{x}} = argmin_x \frac{1}{2}\left\|A(\boldsymbol{x}) - \tilde{\boldsymbol{k}}\right\|^2 + \lambda\psi(\boldsymbol{x}), \tag{2}$$

$$\boldsymbol{x}^{t+1} = \boldsymbol{x}^t - \eta^t\left(A^*\left(A(\boldsymbol{x}) - \tilde{\boldsymbol{k}}\right) + \lambda\phi\left(\boldsymbol{x}^t\right)\right). \tag{3}$$

Here, A is a linear operator that applies sensitivity maps, performs a 2D Fourier transform, and then under-samples the k-space. A^* is the Hermitian of A, ψ is a regularization term, and ϕ is the gradient of ψ with respect to $\boldsymbol{x}$. η^t is the learning rate.

The variational network uses a small convolutional neural network (CNN) for each gradient step in Eq. (3),

$$\boldsymbol{x}^{t+1} = \boldsymbol{x}^t - \eta^t\left(A^*\left(A(\boldsymbol{x}) - \tilde{\boldsymbol{k}}\right) + CNN\left(\boldsymbol{x}^t\right)\right). \tag{4}$$

The end-to-end variational network (E2E-VarNet)[23] shows that the aforementioned gradient update can be further formulated as

$$\boldsymbol{k}^{t+1} = \boldsymbol{k}^t - \eta^t M\left(\boldsymbol{k}^t - \tilde{\boldsymbol{k}}\right) + G(\boldsymbol{k}^t), \tag{5}$$

where $G\left(\boldsymbol{k}^t\right) = \mathcal{F} \circ \mathcal{E} \circ CNN(\mathcal{R} \circ \mathcal{F}^{-1}(\boldsymbol{k}^t))$. $\mathcal{E}$ and $\mathcal{R}$ are the expand and reduce operators. The expand operator takes images and sensitivity maps S_i and outputs corresponding images, while the reduce operator combines individual coil images. $\mathcal{E}(\boldsymbol{x}) = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = (S_1\boldsymbol{x}, \ldots, S_N\boldsymbol{x})$ and $\mathcal{R}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sum_i S_i^*\boldsymbol{x}_i$. For a variational network, the cascading CNN in each step can be a small U-Net structure.[9]

B. Dual-domain learning

Compared to one network performing reconstruction only in the k-space domain or image domain, dual-domain learning has shown overall better image qualities.[24] A dual-domain learning structure includes a k-space reconstruction network and an image domain reconstruction network linked by an inverse Fourier transform layer. In our setup, we take advantage of the superior performance of the variational network and have a simple U-Net for further image domain refinement. Our network has a balanced weight in both domains so that each sub-network gets properly trained.

C. Spatial- and channel-wise attention

For U-Nets in both domains, we adopted spatial and channel wise attention mechanisms.[14] Suppose the feature map for a certain intermediate CNN block has the shape: $X \in \mathbb{R}^{H\times W\times C}$, where $H \times W$ is the shape of the map and C is the number of channels of the feature map. For k-space input, we treated the real part and the imaginary part as two channels stacked, where each channel has real values inside.

The feature map can be seen as a combination of all channels: $X = [x_1, x_2, \ldots, x_C]$, $x_i \in \mathbb{R}^{H\times W}$. Channel-wise attention is achieved by first squeezing the spatial dimensions: taking a global average for each feature map,

$$z_k = \frac{1}{H \times W}\sum_{i}^{H}\sum_{j}^{W} x_k(i,j). \tag{6}$$

Vector $z_k \in \mathbb{R}^{1\times 1\times C}$ contains the global spatial information of the feature map. It is then passed to a fully connected network, followed by a sigmoid activation function to learn the channel-wise attention,

$$s_c = \sigma\big(W_{fc}(z)\big). \tag{7}$$

Here, we used the ReLU function as the intermediate activation for the fully connected network $W_{fc} \in \mathbb{R}^{C\times \frac{C}{2}\times C}$. X was element-wisely multiplied by s_c as the output,

$$y_c = s_c \odot X. \tag{8}$$

Spatial-wise attention tries to learn a score for each pixel considering all channels. Indexing the same feature map spatially, we have $X = [x_{1,1}, x_{1,2}, \ldots, x_{i,j}, \ldots, x_{H,W}]$, $x_{i,j} \in \mathbb{R}^{1\times 1\times C}$. The spatial-wise attention is learned by applying a pixel-wise convolution $W_{conv} \in \mathbb{R}^{1\times 1\times C\times 1}$, followed by a sigmoid function,

$$s_s = \sigma(W_{conv} \circledast X), \tag{9}$$

where $s_s \in \mathbb{R}^{H\times W\times 1}$. In addition, the output is the element-wise multiplication of s_s and X,

$$y_s = s_s \odot X. \tag{10}$$

As shown above, the channel-wise attention squeezes the spatial dimension and learns an important factor for each channel, while the spatial-wise attention squeezes the channel dimension and learns an important factor for each pixel. The final output is the element-wise max-out of these two types of attention,

$$y = \max(y_c, y_s). \tag{11}$$

D. Attention hybrid variational network

Our attention hybrid variational network (AttHybrid-VarNet) structure is shown in Fig. 1. It consists of an end-to-end VarNet (E2E-VarNet) for the k-space domain learning and a refinement network for the image domain learning. Similar to the E2E-VarNet, our AttHybrid-VarNet uses k-space quantities rather than image-space quantities as model input only and requires no pre-training, fine-tuning, or parameter freezing process, making it an end-to-end model. We use a weighted combination of normalized root mean squared errors (NRMSE) for both the k-space domain and image domain learning,

$$L = NRMSE(x, \hat{x}_{intermediate}) + \alpha NRMSE(x, \hat{x}), \tag{12}$$

$$NRMSE(x, \hat{x}) = \frac{\sqrt{MSE(x, \hat{x})}}{\max(x) - \min(x)}, \tag{13}$$

where $\hat{x}_{intermediate}$ and $\hat{x}$ are the intermediate and final reconstructions of the fully sampled reconstruction image x. α is a weighting factor for the image-domain refinement. In our setup, we set $\alpha = 1$ for balanced dual-domain learning.

E. Network, training and dataset details

As shown in Fig. 1, the k-space domain network first takes multi-channel k-space data as input and the sensitivity map estimation (SME) module then gives out estimated sensitivity maps. A series of cascading data consistency (DC) and refinement (R) modules perform reconstruction and give out estimated multi-coil k-space $\hat{\boldsymbol{k}}_i$, where i indicates the i-th coil. These multi-channel k-space data are converted to the image space and combined by applying an inverse Fourier transformation with a root-sum-squares (RSS) at the pixel level for an intermediate reconstruction image,

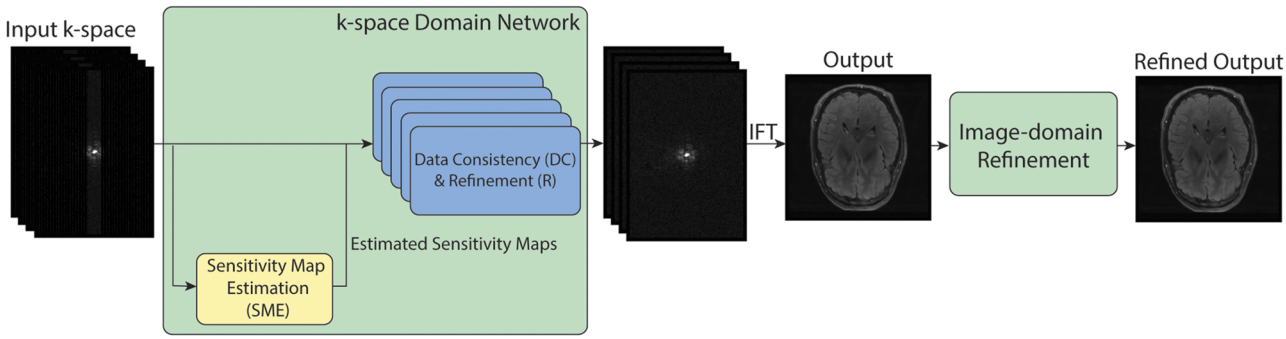

FIG. 1. Overall structure of our attention hybrid variational network (AttHybrid-VarNet). It consists of a variational network for k-space domain learning and an image domain refinement network.

$$\widehat{\boldsymbol{x}}_{intermediate} = RSS(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \sqrt{\sum_{i=1}^{N} |\boldsymbol{x}_i|^2}, \tag{14}$$

where $\boldsymbol{x_i} = \mathcal{F}^{-1}(\widehat{\boldsymbol{k}_i})$. The image domain network takes the intermediate combined image from the multi-coil data as input and performs further refinement. The image domain refinement network design can be rather flexible. It can be a single CNN-based network or even an unrolled structure cascading multiple networks for step-by-step refinement. In our setup, we used a U-Net structure for simplicity. We followed the structure proposed in Ref. 9 with 32 channels for the shallowest layer's convolutional block.

In addition, we applied an attention layer at the end of each depth's convolution block for the U-Net in the image domain refinement network and the k-space domain VarNet. Figure 2 illustrates the structure of the attention layer. We compared our AttHybrid-VarNet against U-Net, W-Net,[24] and E2E-VarNet. We used PyTorch[25] to implement our network. During the training process, we used the Adam optimizer[26] with a learning rate of 0.001 for all the models mentioned.

We evaluated our model on a large-scale open-source MRI dataset (fastMRI) and a dataset derived from patients imaged at our tertiary care hospital and diagnosed with stroke. For the fastMRI dataset, we used the raw k-space data during training; the coil numbers and heights vary case-by-case and have the same width number of 320. During testing, each slice was center cropped to 320×320 for evaluation. We ran tests on both the knee and brain fourfold and eightfold reconstruction tasks, with a center fraction of 0.08 in the fourfold and 0.04 in the eightfold.[11] Due to its large data scale, we randomly selected 521 and 131 cases from the brain dataset for training and testing. For the knee dataset, we used the full single-coil training and validation set for training and testing.

In addition to the fastMRI dataset, we employed a retrospective clinical dataset of patients diagnosed with stroke at our institution. Our dataset included MRI brain scans from patients performed at an urban tertiary referral academic medical center that is a comprehensive stroke center. Clinical scans were conducted between January 1, 2013 and January 1, 2021, on adult patients aged 18–89 years old with recent (acute or subacute) strokes identified for inclusion in this study via a search of the Philips Performance Bridge. Scans meeting this criterion were downloaded and anonymized simultaneously in order to preserve patient anonymity and prevent disclosure of protected health information as part of this IRB exempt study. No patient demographic information was retained for the scans, as it was considered to represent an unnecessary risk for the accidental release of protected health information. Our dataset was collected from both 1.5T and 3T MRI scanners (Philips Healthcare), including 243 patients. In all cases, the b1000 images from a diffusion weighted sequence (DWI) were employed for analysis and acquired using a 2D acquisition with a slice thickness of 5 mm. The acquisition parameters for the 1.5T scans are as follows: TE/TR = 68.8 ms/4183 ms, $FOV = 240 \times 240$ mm^2, and *pixel size* $= 0.94 \times 0.94$ mm^2. For 3T scans: TE/TR = 86.0 ms/4105 ms, $FOV = 230 \times 230$ mm^2, and *pixel size* $= 1.2 \times 1.2$ mm^2. Our b1000 dataset included 7389 slices in total, with 1650 slices containing strokes. We randomly split it using the ratio of 80%/20%, leading to 5904 slices for training and 1485 slices for testing. We took measurements in the MR imaging using 2-D acquisition and slice-level normalization to avoid the dependency among slices from the same subject to a great extent. The b1000 dataset contains multiple image sizes, with the MR image as the raw data. We first resized all image slices to 256×256 and applied a Fourier transformation to get the corresponding k-space data. We performed more challenging 20-fold and 30-fold acceleration reconstruction tasks with two-dimensional Gaussian sampling.

To evaluate the image quality of our reconstructions, we report peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) with respect to fully sampled ones. Furthermore, we conducted a blinded image quality test using a single board-certified radiologist with subspecialty certification in neuroradiology (C.W.F.) for evaluation of the quality of the reconstructions of four models. We had the qualified reader review randomly selected slices from the different models to see how qualitatively reviewed images by an expert neuroradiologist compared in addition to our quantitative metrics, as the visual appeal of images to the expert reader may impact the

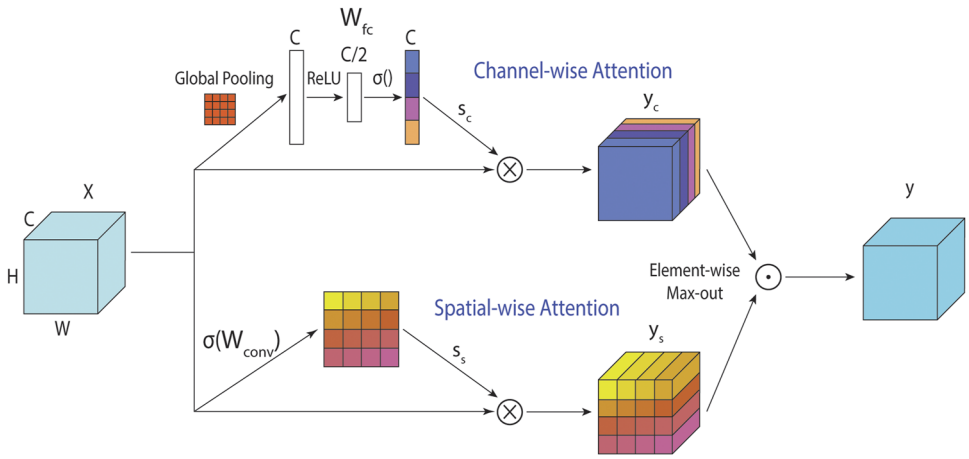

FIG. 2. Detailed structure of the attention layer. It includes two attention paths: channel-wise attention and spatial-wise attention. The outputs y_c and y_s from two paths are followed by an element-wise max-out operation to get the final output y.

utility of the utilization of these images in a clinical setting. We randomly extracted 100 slices of 79 patients with adequate brain area from the test set. For each slice sample, the radiologist was given four candidate reconstructions and asked to rank their preference from 1 (most preferable) to 4 (least preferable). The priority score is then given by

$$S_{priority} = \frac{N_{candidates} + 1 - \frac{1}{N}\sum_{i=1}^{N} p_i}{N_{candidates}}, \tag{15}$$

where $N_{candidates} = 4$ is the number of candidates for each slice being evaluated. N is the number of samples (slices). p_i stands for the

TABLE I. Summary of results on the fastMRI dataset for different models over multiple acceleration factors and reconstruction tasks. Boldface denotes superior performance for the corresponding index.

Brain				
	4×		8×	
	PSNR	SSIM	PSNR	SSIM
U-Net	38.08 ± 0.26	0.9458 ± 0.0027	34.10 ± 0.28	0.9176 ± 0.0035
W-Net	39.52 ± 0.29	0.9505 ± 0.0028	34.87 ± 0.30	0.9200 ± 0.0038
E2E-VarNet	40.81 ± 0.29	0.9568 ± 0.0026	36.75 ± 0.25	0.9340 ± 0.0030
AttHybrid-VarNet	**40.92 ± 0.29**	**0.9577 ± 0.0025**	**37.03 ± 0.25**	**0.9365 ± 0.0029**
Knee				
	4×		8×	
U-Net	30.45 ± 0.23	0.6777 ± 0.0091	28.55 ± 0.19	0.6038 ± 0.0102
W-Net	30.61 ± 0.23	0.6808 ± 0.0091	28.73 ± 0.20	0.6060 ± 0.0103
E2E-Varnet	31.07 ± 0.26	0.6899 ± 0.0095	29.48 ± 0.22	0.6187 ± 0.0107
AttHybrid-VarNet	**31.09 ± 0.25**	**0.6901 ± 0.0094**	**29.49 ± 0.22**	**0.6197 ± 0.0106**

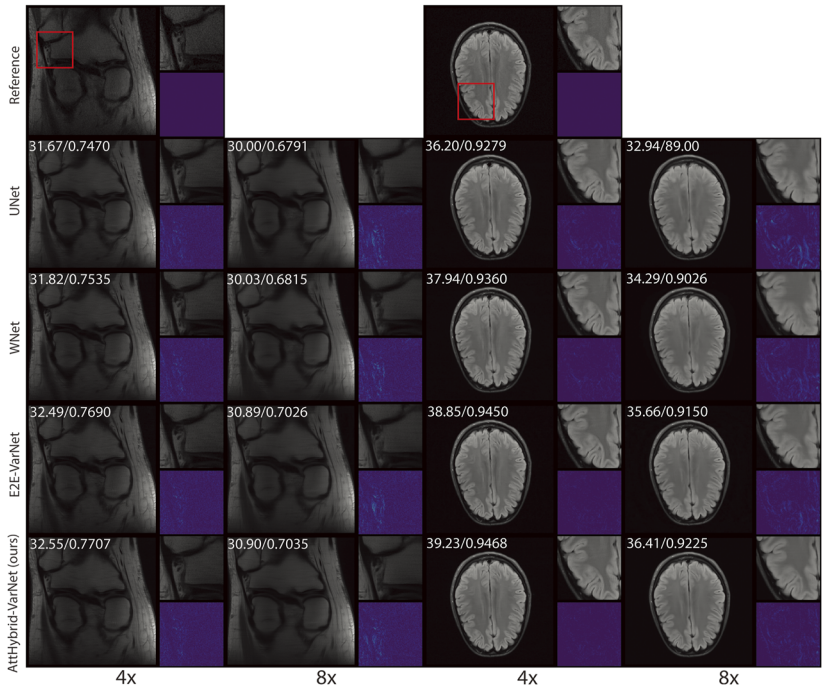

FIG. 3. Reconstruction samples from the fastMRI dataset. Red boxes in the fully sampled reference scans highlight areas for the enlarged patches and the corresponding error maps. The numbers on the upper left of each image indicate PSNR and SSIM, respectively.

priority rank. Higher priority score $S_{priority}$ means it is more preferred in the blind test.

III. RESULTS

A. Performance on fastMRI dataset

The fastMRI brain dataset contains 8336 slices for training and 2096 slices for testing. Table I demonstrates the evaluation metrics. For the fourfold acceleration task in brain reconstruction, our model achieved an overall *PSNR* = 40.92 and an *SSIM* = 0.9577. In the eightfold brain reconstruction task, our model achieved an overall *PSNR* = 37.03 and an *SSIM* = 0.9365. As for the knee reconstruction tasks, for fourfold acceleration, our model had an overall *PSNR* = 31.09 and an *SSIM* = 0.6901. Moreover, *PSNR* = 29.49 and *SSIM* = 0.6197 for the eightfold acceleration test. Reconstruction samples for the fastMRI dataset are shown in Fig. 3. The size of the

TABLE II. Summary of results on the b1000 dataset for different models over multiple acceleration factors and reconstruction tasks. Boldface denotes superior performance for the corresponding index.

	PSNR	SSIM
	20×	
U-Net	29.67 ± 0.13	0.8230 ± 0.0034
W-Net	31.73 ± 0.13	0.8434 ± 0.0037
E2E-VarNet	36.25 ± 0.16	0.9039 ± 0.0030
AttHybrid-VarNet	**36.32 ± 0.16**	**0.9199 ± 0.0029**
	30×	
U-Net	28.45 ± 0.12	0.7920 ± 0.0040
W-Net	30.50 ± 0.13	0.8193 ± 0.0040
E2E-VarNet	33.37 ± 0.15	0.8642 ± 0.0034
AttHybrid-VarNet	**33.70 ± 0.15**	**0.8882 ± 0.0035**

TABLE III. Raw summed priority rank and priority score for b1000 reconstructions over multiple models evaluated by radiologists. Boldface denotes superior performance for the corresponding index.

	Raw priority rank	
	20×	30×
U-Net	396	396
W-Net	296	301
E2E-VarNet	164	161
AttHybrid-VarNet	**144**	**142**
	Priority score	
U-Net	0.26	0.26
W-Net	0.51	0.50
E2E-VarNet	0.84	0.85
AttHybrid-VarNet	**0.89**	**0.90**

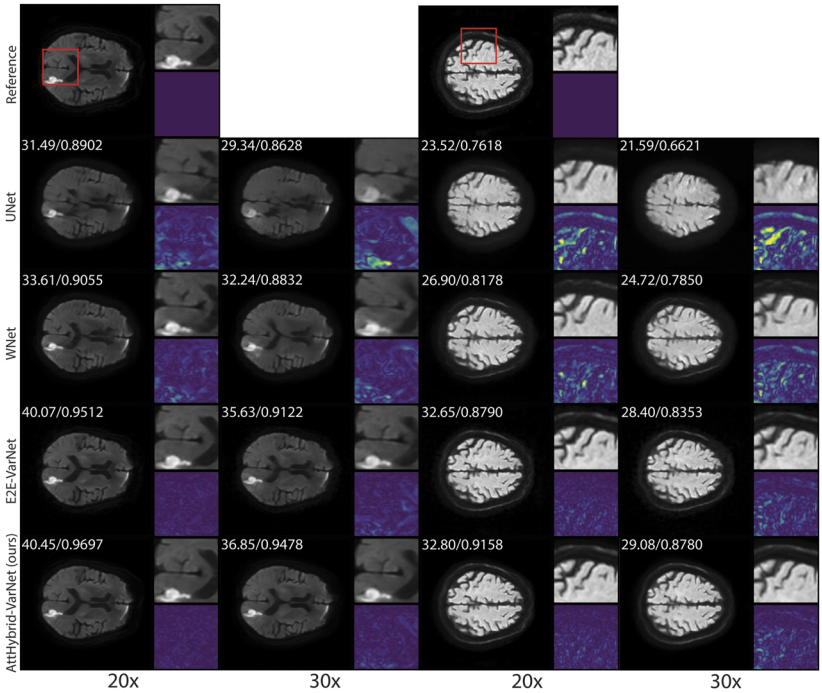

FIG. 4. Reconstruction samples from the b1000 dataset. Red boxes in the fully sampled reference scans highlight areas for the enlarged patches and the corresponding error maps. The numbers on the upper left of each image indicate PSNR and SSIM, respectively.

image is 320 × 320 pixels. Red boxes depict windows with a size of 90 × 90 pixels, with their enlarged details shown on the upper-right side and their corresponding error maps on the lower-right side. All the error maps are unit normalized and enhanced three times for better demonstration. All models performed well in the four-fold test, as the acceleration factor was small. For the eightfold test, our model yielded cleaner error maps for both the brain and knee reconstruction. Interestingly, although E2E-VarNet achieved better evaluation metrics than W-Net in the eightfold tasks, their enlarged area depicted a similar level of error.

B. Performance on b1000 dataset

Table II illustrates the evaluation metrics of the b1000 dataset. Our dataset included 5904 slices for training and 1485 slices for testing. To demonstrate the effectiveness of our model, we ran 20- and 30-fold acceleration reconstructions for the b1000 dataset. For the 20-fold acceleration, our model achieved an overall *PSNR* = 36.32 and an *SSIM* = 0.9199. For 30-fold tasks, our model achieved an overall *PSNR* = 33.70 and an *SSIM* = 0.8882.

Table III shows the blinded image quality test results for image quality preference over 100 sample slices from the b1000 dataset. The raw priority rank was calculated by summing up 100 priority ranks for each candidate model as determined by the radiologist. Then, the priority score was calculated according to Eq. (15). Higher priority scores represent the more preferred reconstruction from the corresponding candidate model choices. Our model achieved an overall score of 0.89 and 0.90 for the 20- and 30-fold tests, respectively.

Figure 4 depicts sample reconstructions for the b1000 dataset. The size of the image is 256 × 256 pixels, with red boxes highlighting a window of size 70 × 70 pixels. Enlarged image details and their corresponding error maps are shown on the right side of the image. All the error maps are unit normalized and enhanced three times for better demonstration. The improvement from U-Net to E2E-VarNet is even more noticeable under higher acceleration factors thanks to the unrolled optimization structure. The comparison between the W-Net and U-Net illustrates the benefits of performing dual-domain learning. Our model benefits from both and gives cleaner error maps in samples with or without stroke.

IV. CONCLUSION

Prior studies on accelerated MRI reconstruction have typically focused on single-domain learning. Previous dual-domain learning structures, such as W-Net, have a similar structure modality in both domains. In contradistinction, in this study, we focused on building a dual-domain learning structure with different modalities in each domain. Unrolled optimization structures, such as variational networks, have been reported to be more suitable for k-space reconstruction,[22,23] which inspired our use of an attention hybrid variational network. Notably, the choice for the image domain refinement network can be flexible, and with larger and finer structures, one can anticipate further improvements in performance. Furthermore, one also needs to consider the training process according to the structure design. Larger refinement networks in the image domain can lead to a longer and more challenging training process. This can be mitigated, however, by using fine-tuning with an out-of-box pre-trained model checkpoint. Nevertheless, we kept the structure in the image-domain as simple as possible such that our model can be trained in an end-to-end manner.

Apart from the widely used fastMRI dataset, we tested our model on our stroke dataset and performed an evaluation by a single subspecialty trained radiologist. Rather than using the small acceleration factors recommended by the fastMRI dataset, we tested our model with more challenging cases to demonstrate its effectiveness. When compared to the results from the fastMRI dataset, the difference among models is even more noticeable. In fact, the priority score of the E2E-VarNet when compared to U-Net and W-Net further validates the effectiveness of a variational net for k-space learning. Ultimately, our model demonstrated an overall superior performance in numerical metrics and blinded image analysis.

Several limitations of our study are of note. First, as an image regression model, smoothness in the reconstructed images was an expected effect, which is inherited in the loss function and training procedure. In fact, loss functions such as mean squared error, absolute value error, or structural similarity error all lead to smoothness. This can be mitigated by dithering the image with a small amount of Gaussian noise in order to preserve the sharpness. More delicate loss functions and training processes would be an interesting topic for future consideration. Second, similar to previous models, our model is task specific. One large topic for future consideration would be designing and training a universal reconstruction network. Recent developments in generative AI have shown promising results in the common image contents field. One challenge in developing a universal reconstruction model would be the requirement for an even larger dataset. Another related question would relate to the fashion by which to properly encode the reconstruction task for a universal model. Nevertheless, this avenue of inquiry would be an interesting topic for future development efforts.

In summary, we report the development of an attention hybrid variational network for accelerated MRI reconstruction. Our model benefits from an unrolled optimization structure and dual-domain learning. We tested our model on a large-scale dataset and then validated our model on a clinically relevant stroke database from our own institution. We performed numerical evaluation and blinded image quality analyses to demonstrate the effectiveness of our model. In future studies, we hope this work can serve as a reference for cross-domain multi-modality networks for image reconstruction.

ACKNOWLEDGMENTS

This work was supported by the Rajen Kilachand Fund for Integrated Life Science and Engineering. We would like to acknowledge the Boston University Photonics Center for technical support. In addition, G.S. acknowledges the BUnano Cross-Disciplinary Fellowship from Boston University.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

No animal or human experiments are included in this work.

Author Contributions

Guoyao Shen: Conceptualization (equal); Formal analysis (lead); Methodology (lead); Software (lead); Writing – original draft (lead). **Boran Hao**: Methodology (supporting); Software (supporting). **Mengyu Li**: Software (supporting). **Chad W. Farris**: Formal analysis (supporting); Writing – review & editing (supporting). **Ioannis Ch. Paschalidis**: Writing – review & editing (supporting). **Stephan W. Anderson**: Conceptualization (equal); Methodology (supporting); Writing – review & editing (equal). **Xin Zhang**: Conceptualization (equal); Funding acquisition (lead); Project administration (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The fastMRI dataset that supports the findings of this study is openly available at https://fastmri.med.nyu.edu/.[27] The codes of this study are openly available at https://github.com/GuoyaoShen/AttHybrid-VarNet.[28]

REFERENCES

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," Magn. Reson. Med. **58**(6), 1182–1195 (2007).

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory **52**(2), 489–509 (2006).

[3] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Artif. Intell. Rev. **53**(8), 5455–5516 (2020).

[4] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, "Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI," J. Magn. Reson. Imaging **49**(4), 939–954 (2019).

[5] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," Z. Med. Phys. **29**(2), 102–127 (2019).

[6] J. Montalt-Tordera, V. Muthurangu, A. Hauptmann, and J. A. Steeden, "Machine learning in Magnetic Resonance Imaging: Image reconstruction," Phys. Med. **83**, 79–87 (2021).

[7] A. Pal and Y. Rathi, "A review and experimental evaluation of deep learning methods for MRI reconstruction," Mach. Learn. Biomed. Imaging **1**, 1 (2022).

[8] S. Geethanath, R. Reddy, A. S. Konar, S. Imam, R. Sundaresan, R. Babu, and R. Venkatesan, "Compressed sensing MRI: A review," Critical Rev. Biomed. Eng. **41**(3), 183–204 (2013).

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention: MICCAI* 2015 (IEEE, 2015), pp. 234–241.

[10] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," IEEE Trans. Image Process. **26**(9), 4509–4522 (2017).

[11] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, N. Yakubova, J. Pinkerton, D. Wang, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui, "fastMRI: An open dataset and benchmarks for accelerated MRI," arXiv:1811.08839 (2019).

[12] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," IEEE Trans. Med. Imaging **37**(2), 491–503 (2018).

[13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[14] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating fully convolutional networks with spatial and channel 'squeeze and excitation' blocks," IEEE Trans. Med. Imaging **38**(2), 540–549 (2019).

[15] M. H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," Comput. Visual Media **8**(3), 331–368 (2022).

[16] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," Neurocomputing **452**, 48–62 (2021).

[17] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021* (IEEE, 2021), pp. 13713–13722.

[18] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," Nature **555**(7697), 487–492 (2018).

[19] T. Eo, H. Shin, Y. Jun, T. Kim, and D. Hwang, "Accelerating Cartesian MRI by domain-transform manifold learning in phase-encoding direction," Med. Image Anal. **63**, 101689 (2020).

[20] K. Liang, H. Yang, and Y. Xing, "Comparison of projection domain, image domain, and comprehensive deep learning for sparse-view X-ray CT image reconstruction," arXiv:1804.04289 (2019).

[21] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep MRI reconstruction: Unrolled optimization algorithms meet neural networks," arXiv:1907.11711 (2019).

[22] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," Magn. Reson. Med. **79**(6), 3055–3071 (2018).

[23] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson, "End-to-end variational networks for accelerated MRI reconstruction," in *Medical Image Computing and Computer Assisted Intervention: MICCAI 2020* (IEEE, 2020), pp. 64–73.

[24] R. Souza and R. Frayne, "A hybrid frequency-domain/image-domain deep network for magnetic resonance image reconstruction," in *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Rio de Janeiro, Brazil* (IEEE, 2019), pp. 257–264.

[25] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 32 (MIT Press, 2019).

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2017).

[27] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, N. Yakubova, J. Pinkerton, D. Wang, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui (2019). "fastMRI: An open dataset and benchmarks for accelerated MRI," fastMRI Dataset. https://fastmri.med.nyu.edu/

[28] G. Shen (2023). "Attention hybrid variational net for accelerated MRI reconstruction," GitHub. https://github.com/GuoyaoShen/AttHybrid-VarNet