IR-FRestormer: Iterative Refinement with Fourier-Based Restormer for Accelerated MRI Reconstruction

Mohammad Zalbagi Darestani*
Rice University
mz35@rice.edu

Vishwesh Nath
NVIDIA
vnath@nvidia.com

Wenqi Li
NVIDIA
wenqil@nvidia.com

Yufan He
NVIDIA
yufanh@nvidia.com

Holger R. Roth
NVIDIA
hroth@nvidia.com

Ziyue Xu
NVIDIA
ziyuex@nvidia.com

Daguang Xu
NVIDIA
daguangx@nvidia.com

Reinhard Heckel
Technical University of Munich
reinhard.heckel@tum.de

Can Zhao
NVIDIA
canz@nvidia.com

Abstract

Accelerated magnetic resonance imaging (MRI) aims to reconstruct high-quality MR images from a set of under-sampled measurements. State-of-the-art methods for this task use deep learning, which offers high reconstruction accuracy and fast runtimes. In this work, we propose a new state-of-the-art reconstruction model for accelerated MRI reconstruction. Our model is the first to combine the power of deep neural networks with iterative refinement for this task. For the neural network component of our method, we utilize a transformer-based architecture as transformers are state-of-the-art in various image reconstruction tasks. However, a major drawback of transformers which has limited their emergence among the state-of-the-art MRI models is that they are often memory inefficient for high-resolution inputs. To address this limitation, we propose a transformer-based model which uses parameter-free Fourier-based attention modules, achieving 2× more memory efficiency. We evaluate our model on the largest publicly available MRI dataset, the fastMRI dataset [46], and achieve on-par performance with other state-of-the-art[1] methods on the dataset's leaderboard[2].

*The work was mainly done during an internship at Nvidia Research.

[1]At the time of writing this manuscript, the best-performing model on the leaderboard is AIRS-Net whose implementation is not published. Our implementation, based on the limited material available for AIRS-Net, does not reproduce the results of the leaderboard.

[2]https://fastmri.org/leaderboards/

1. Introduction

Magnetic resonance imaging (MRI) is a crucial medical imaging modality that offers non-invasive, ionizing radiation-free visualization of soft tissues. However, acquiring a fully-sampled MRI scan requires a long acquisition time, making it infeasible in most clinical practice. To address this challenge, accelerated MRI is widely used to accelerate the acquisition process via recording less measurements in the Fourier space (also called k-space) than a fully-sampled scan. The process of reconstructing the MR image from the under-sampled k-space data is called Accelerated MRI reconstruction.

Accelerated MRI reconstruction has a rich history in the compressed sensing field, and traditional compressed sensing methods such as sparsity-based methods have been successfully applied to this task [5, 26]. However, these methods have major drawbacks, including (1) being extremely slow at inference, (2) not being able to achieve a high and clinically-accepted reconstruction quality, and (3) being difficult to tune for achieving high reconstruction accuracy on multiple samples given a fixed set of hyper-parameters.

The limitations of traditional methods have motivated the development of new techniques, particularly deep learning-based approaches. With the emergence of deep learning, convolutional neural networks (CNNs) have demonstrated superior performance over classical methods both in terms of accuracy and runtime [39, 10].

More recently, the fastMRI dataset [46] was released as the largest publicly available dataset for deep learning based accelerated MRI reconstruction. The fastMRI

dataset enabled deep learning models to compete for advancing state-of-the-art [19, 27]. The fastMRI dataset also has withheld a test set for evaluation purposes, the results of which are reported to a public leaderboard, to keep track of the state-of-the-art models in the field. The leaderboard is divided into two general tracks according to the anatomy of scans (brain or knee). Furthermore, each track is divided into two tracks for 4×- or 8×-accelerated MRI reconstruction.

In this work, we propose a novel model for the brain 8× track of the fastMRI leaderboard. We target this track as the most challenging one since it has been advanced only once over the past two years, and it also deals with the reconstruction from 8×-accelerated measurements, which is more difficult than 4× acceleration. Our model, which is based on iterative refinement via an efficient transformer-based network, achieves competitive results with much fewer number of parameters. To summarize, our key contributions are as follows:

- We propose a novel MR image reconstruction algorithm based on iterative refinement with neural networks. Our work is the first to employ iterative refinement for MRI reconstruction.

- We propose a new transformer network called FRestormer, which is specifically designed to address the memory issue of transformers for high-resolution medical imaging. This modification also enables embedding the network inside our iterative refinement paradigm.

- We demonstrate that our model achieves on-par performance with state-of-the-art models on the fastMRI [46] leaderboard. We provide quantitative as well as qualitative comparisons between our method and other state-of-the-art models.

- We are the first transformer-based method to achieve competitive performance on the fastMRI leaderboard for the brain track, in line with top-performing methods.

- Contrary to state-of-the-art unrolled networks, our method does not rely on unrolling an optimization algorithm. Rather, the network is used in a closed loop where its output is refined and fed to the network multiple times to generate the result. Thus, a competitive advantage of our framework is that one may use a model of higher capacity contrary to unrolled networks where multiple small networks are attached sequentially to generate the output.

1.1. Problem setup

Accelerated MRI reconstruction is the problem of reconstructing an image $\mathbf{x}^* \in \mathbb{C}^N$ from under-sampled measurements (also called under-sampled k-space)

$$\mathbf{y}_i = \mathbf{MFS}_i\mathbf{x}^* + \text{noise} \in \mathbb{C}^{M_c}, \quad i = 1, \dots, n_c. \tag{1}$$

Here, n_c denotes the number of radiofrequency coils, $\mathbf{S}_i$ is a complex-valued position-dependent coil sensitivity map, that is applied through element-wise multiplication to the image $\mathbf{x}^*$, $\mathbf{F}$ is the 2D discrete Fourier transform, and $\mathbf{M}$ is a mask (a diagonal matrix with ones and zeros on its diagonal) that implements under-sampling of k-space data. The forward operator $\mathbf{A} = \mathbf{MFS}$ is known a priori. The problem setup above is for the accelerated multi-coil MRI reconstruction which is more common in practice than single-coil where $n_c = 1$. Note that should a reconstruction algorithm use $\mathbf{A}$ for multi-coil MRI reconstruction (e.g., [34]), it needs to have access to an estimate of coil sensitivities, otherwise no information of coil sensitivity maps is needed.

1.2. Accelerated MRI reconstruction with neural networks

For accelerated MRI reconstruction, neural networks can be used via (1) supervised learning [1, 34], (2) unsupervised learning [37, 15, 7, 2], and (3) self-supervised learning [42]. The network typically works by mapping either the under-sampled k-space to the high-quality artifact-free image, or by mapping the coarse image reconstructed from the under-sampled measurement (e.g., the zero-filled image) to the high-quality image.

As the focus of this work, *supervised* methods train a network on a dataset of pairs of under-sampled measurement and ground-truth image $\{(\mathbf{x}_1^*, \mathbf{y}_1), \dots, (\mathbf{x}_n^*, \mathbf{y}_n)\}$. The training loss (e.g., the ℓ_1 loss) is then computed between network output and the ground-truth image. Supervised models are lightning-fast at inference as one needs to only make a forward pass through the network to obtain the reconstructed image. State-of-the-art MRI reconstruction methods, including ours, are all supervised according to the fastMRI leaderboard [46].

1.3. Related work

Over the past few years, several families of neural network architectures have been applied to accelerated MRI reconstruction. In the following, we iterate through several of such architectural categories:

CNNs: One of the first deep learning based model that was applied to accelerated MRI reconstruction was a simple CNN with several convolutional layers [39].

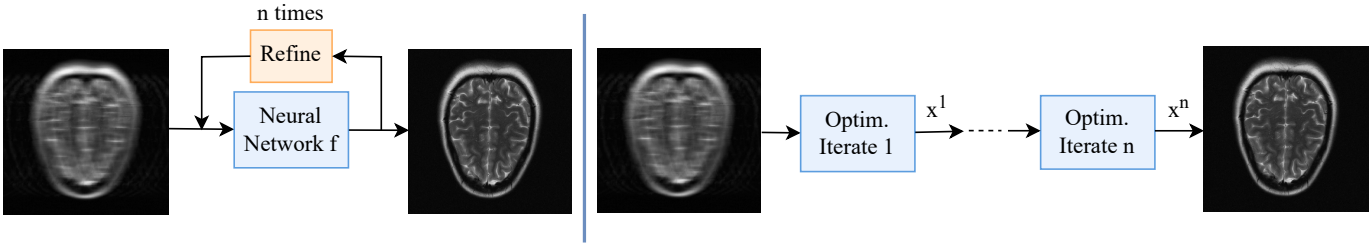

Figure 1: **Left:** Our neural-network-based iterative refinement framework. Our network repeatedly refines its output n times. One step i of our method contains two parts: (1) *reconstruction* $\mathbf{x}^i = f(\mathbf{x}^i)$ which is performed by the *fourier* neural network f, and (2) *refinement* $\mathbf{x}^{i+1} = \text{Prox}(\mathbf{x}^i) = \mathbf{x}^i - (\mathbf{A} + \frac{1}{\eta}\mathbf{I})^{-1}(\mathbf{A}\mathbf{x}^i - \mathbf{y})$ which is done via the proximal operator of proximal gradient descent ($\mathbf{A}$ is the forward operator, $\mathbf{I}$ is the identity matrix, and $\mathbf{y}$ is the under-sampled k-space). **Right:** Conventional algorithmic unrolling where several optimization iterates are modeled via neural networks and trained end to end (this approach is the current state of the art).

Later on, other CNN architectures like the U-Net [32] and the KIKI-Net [9] were adopted for reconstruction and achieved higher reconstruction performance than a vanilla CNN [17]. The main characteristic of CNNs for image reconstruction is their ability to model and extract local features in an image via convolutional filters. Thus, unawareness of global structures in the image is the main limitation of CNN-based architectures.

Generative models: Generative models are successful tools for various image synthesis applications. They have also been applied to MRI reconstruction [18, 16, 33, 44]. However, they fall behind state-of-the-art MRI reconstruction models in terms of reconstruction accuracy and runtime. For instance, there is no generative model among the top-performing models on the fastMRI leaderboard. Their long inference time also makes them impractical for clinical deployment.

Transformers: Transformers, originally proposed for natural language processing (NLP) [38], have entered the computer vision field through the Vision Transformer (ViT) [8] model. ViT has been tailored to accelerated MRI reconstruction as well [12, 13, 21, 20], and has been shown to be computationally faster than U-Net [23], but its reconstruction accuracy is only marginally better than the U-Net. The main benefit attributed to transformers is their ability to model long-range dependencies across the image. This is useful when redundant information is distributed all over the image and one does not observe dependencies only locally (for example, the white matter appears in many places of a brain MRI image with a similar texture). However, transformers have had limited success on the fastMRI leaderboard, with only one out of nine top-performing models based on a transformer architecture. This is most likely due to their memory inefficiency in handling high-resolution images of the fastMRI dataset. The reason why the only top-performing transformer-based method has not been used on the brain track might include that brain images have a larger size and require more GPU memory than knee images. To overcome this, we propose an architecture modification to make transformers more memory efficient and leverage their superior performance for our method.

Unrolled networks: These models are rooted in optimization methods. Since this family of methods are closely related to our approach, let us introduce this class with an elaborate example. In order to reconstruct the ground-truth image, suppose we run gradient descent on the following optimization problem:

$$\arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \Psi(\mathbf{x}). \tag{2}$$

Here, $\Psi(\mathbf{x})$ represents an explicit regularization function imposing a prior on $\mathbf{x}$ (e.g., it could be $\|\mathbf{x}\|_{\text{TV}}$ like classical sparsity-based methods). Thus, the gradient descent update rule at iteration i is:

$$\mathbf{x}^{i+1} = \mathbf{x}^i - \eta\left(\mathbf{A}^T(\mathbf{A}\mathbf{x}^i - \mathbf{y}) + \nabla\Psi(\mathbf{x}^i)\right). \tag{3}$$

The idea of unrolled networks is to approximate $\nabla\Psi(\mathbf{x}^i)$ with a neural network. For example, the end-to-end variational network (E2E-VarNet) [34] uses a U-Net to approximate the gradient of the regularizer. In principle, this approximator can be any neural network f. With this approximation, the update rule changes to:

$$\mathbf{x}^{i+1} = \mathbf{x}^i - \eta\left(\mathbf{A}^T(\mathbf{A}\mathbf{x}^i - \mathbf{y}) + f(\mathbf{x}^i)\right). \tag{4}$$

As a final step, this process is unrolled. Specifically, each step of gradient descent shown in equation (4) is

modeled with a block in the network. Then, a few of these blocks in tandem are trained end-to-end, typically with respect to a supervised loss function.

Among the reconstruction model categories discussed above, unrolled networks are currently state-of-the-art for accelerated MRI reconstruction. Aside from the example above, there are numerous variants of unrolled networks in terms of what type of optimization problem is unrolled [35, 47, 1, 6].

In this work, we propose a new model that is not obtained by unrolling one of the standard optimization algorithms such as gradient descent, but it is very similar to unrolled methods in that it also consists of refinement blocks and data-consistency-enforcing blocks. Our model is based on *iterative refinement*, a concept rooted in numerical methods for solving a linear system of equations [4]. Our iterative refinement method (illustrated in Fig. 1) may be connected to unrolled networks via unrolling a proximal optimization problem as elaborated in Section 2. Our work is the first to use iterative refinement for accelerated MRI reconstruction, although variants of it have been proposed for other applications, e.g., computed tomography (CT) reconstruction [14].

2. Method

In this section, we propose our approach for combining iterative refinement with neural networks. We start with an introduction to iterative refinement and then state our approach to bundle it with a neural network. Afterwards, we elaborate on our choice of neural network architecture (which is a transformer-based network) and our proposed modification.

2.1. Iterative refinement with neural networks

From an optimization perspective, iterative refinement can be viewed as a special case of proximal minimization [29]. Proximal minimization typically deals with solving the following optimization problem:

$$\arg\min_{\mathbf{x}} g(\mathbf{x}) + h(\mathbf{x}), \tag{5}$$

where g and h are two convex functions with different gradient behaviors (for example, g may be differentiable and h may be non-differentiable). Proximal methods make a quadratic approximation to the well-behaved function, say g:

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} g(\mathbf{x}) + h(\mathbf{x}) \\ &\approx \arg\min_{\mathbf{x}} g(\mathbf{z}) + \nabla g(\mathbf{z})^T(\mathbf{x}-\mathbf{z}) + \frac{1}{2\eta}\|\mathbf{x}-\mathbf{z}\|_2^2 + h(\mathbf{x}) \\ &= \arg\min_{\mathbf{x}} \frac{1}{2\eta}\|\mathbf{x}-(\mathbf{z}-\eta\nabla g(\mathbf{z}))\|_2^2 + h(\mathbf{x}).\end{aligned} \tag{6}$$

To solve this optimization problem via proximal gradient descent, one can define a two-step update rule as follows:

$$\begin{aligned}&\text{Step 1 of iterate i}: \quad \mathbf{z} = \mathbf{x}^i - \eta\nabla g(\mathbf{x}^i), \\ &\text{Step 2 of iterate i}: \quad \mathbf{x}^{i+1} = \arg\min_{\mathbf{x}} \frac{1}{2\eta}\|\mathbf{x}-\mathbf{z}\|_2^2 + h(\mathbf{x}).\end{aligned} \tag{7}$$

The second step in the update rule above is also known as the *proximal step* based on which the proximal operator is defined as:

$$\text{Prox}_{h,\eta}(\mathbf{z}) = \arg\min_{\mathbf{x}} \frac{1}{2\eta}\|\mathbf{x}-\mathbf{z}\|_2^2 + h(\mathbf{x}). \tag{8}$$

Iterative refinement is a special case of proximal minimization with $g = 0$ and h being the quadratic function:

$$h(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{y}^T\mathbf{x}. \tag{9}$$

In this case, the update rule in (7) may be written as:

$$\begin{aligned}\mathbf{x}^{i+1} &= \arg\min_{\mathbf{x}} \frac{1}{2\eta}\|\mathbf{x}-\mathbf{x}^i\|_2^2 + h(\mathbf{x}) \\ &= (\mathbf{A} + \frac{1}{\eta}\mathbf{I})^{-1}(\mathbf{y} + \frac{1}{\eta}\mathbf{x}^i) \\ &= \mathbf{x}^i - (\mathbf{A} + \frac{1}{\eta}\mathbf{I})^{-1}(\mathbf{A}\mathbf{x}^i - \mathbf{y}) = \text{Prox}_{h,\eta}(\mathbf{x}^i)\end{aligned} \tag{10}$$

The update rule (10) of iterative refinement is also called the *refinement step* which is done according to the residual $\mathbf{A}\mathbf{x}^i - \mathbf{y}$.

In our approach, **we do not unroll** (10) to learn the regularization term (as there is no explicit regularization term) via a neural network (contrary to the common practice in unrolled neural networks). Instead, we wrap a neural network around the proximal operator $\text{Prox}_{h,\eta}(\mathbf{x}^i)$ and then in an end-to-end manner, we train several iterations of $f\left(\text{Prox}_{h,\eta}(\mathbf{x}^i)\right)$ where f is the neural network (please see Fig. 1 for an illustration of our framework). Intuitively, our method utilizes the *refinement step* of iterative refinement to *refine* the reconstruction estimate given by the neural network.

There is no constraint on the choice of neural network in our framework. However, as we discuss in the following section, a neural network that performs better than another network, also performs better in combination with iterative refinement.

Finally, in order to perform the refinement step, one requires knowledge of the forward operator $\mathbf{A} = \mathbf{MFS}$. The under-sampling mask $\mathbf{M}$ and the Fourier operator $\mathbf{F}$ are known but the coil sensitivity maps $\mathbf{S}$ are not known. However, $\mathbf{S}$ can be computed from

the under-sampled measurements via traditional coil sensitivity estimation methods such as ESPIRiT [36] or SENSE [30]. Alternatively, one can learn them as part of the end-to-end training paradigm which works better than traditional methods [34]. In our work, we follow the modern practice and learn them implicitly using U-Net as part of our framework.

2.2. Choice of neural network architecture

Any neural network can be embedded into our iterative refinement paradigm. We hypothesize that a better performing reconstruction network also performs better when combined with iterative refinement. To confirm this, we compare the baseline U-Net [32] (which is often used as the CNN backbone of modern MRI reconstruction networks [34]) with the state-of-the-art transformer-based image restoration model Restormer [45].

Both networks follow an encoder-decoder like architecture with downsampling and upsampling blocks. Contrary to U-Net which is composed of convolutional blocks in each resolution level of the network, Restormer contains transformer blocks, each containing an attention module and a feed-forward network.

Like many transformer-based models, Restormer is very memory hungry, especially in our work where we deal with extremely high-resolution k-space samples. This becomes practically prohibitive when plugging Restormer to iterative refinement. For example, training iterative refinement with Restormer (5M parameters in total) on the whole multi-coil brain training set of the fastMRI dataset [46] is not feasible on a GPU with 48 GB of memory with more than one refinement step.

To address this issue, we made the following modifications. First, we propose a new transformer network called **FRestormer**. We achieve this by replacing every attention module of Restormer with a simple Fourier transform, as shown in Fig. 2. Previous research has demonstrated that replacing attention modules with non-parametric operators such as Fourier transform or pooling leads to minimal accuracy degradation in NLP tasks [22] and no accuracy drop in image classification [43]. Our ablations (in Section 3.3) demonstrate that Restormer and FRestormer achieve comparable reconstruction accuracy, with statistically insignificant (p>0.5) difference.

Secondly, we cropped samples to a lower size (because fastMRI samples are by default zero-padded in the image domain) for the neural network part of our iterative refinement paradigm (i.e., we worked with the full size in the refinement step). With these modifications, we were able to successfully train iterative refinement with FRestormer, with 3 number of refinement iterates on the whole training set, on 48 GB of GPU memory.

Model	SSIM	p-value
U-Net	0.9282	-
FRestormer	0.9322	<0.01
IR-U-Net	0.9487	-
IR-FRestormer	**0.9503**	<0.01

Table 1: **Our FRestormer outperforms U-Net with and without iterative refinement.** There are 3 iterative steps used in IR-U-Net and IR-FRestormer. Each p-value is the result of paired Wilcoxon signed-rank test between the setup of the row that contains the p-value and its previous row.

Finally, we now compare U-Net with FRestormer in 2 scenarios: (1) when they are trained on their own, and (2) when they are part of our iterative refinement paradigm. For this experiment, we selected a subset of the multi-coil AXT2 brain dataset of fastMRI [46] with the split size of 500/170/140 2D volumes for train/validation/test. For iterative refinement, we used 3 refinement steps. Furthermore, both U-Net and Restormer have 4 encoder and 4 decoder layers with ∼7M and ∼2M parameters, respectively. When iterative refinement is used, 2M additional parameters are used for the coil sensitivity estimation which is done via a U-Net with a width factor of 16.

Table 1 shows the results. We draw two conclusions from this table: (1) FRestormer outperforms U-Net, and (2) IR-FRestormer outperforms IR-U-Net. The second remark confirms our hypothesis that when a network outperforms another one, its combination with iterative refinement also outperforms the combination of the other network with iterative refinement.

3. Experiments

In this section, we start off with our experimental setup, then state our main results, and finally present our ablation studies.

3.1. Experiment setup

Dataset: We worked with the fastMRI dataset containing images that are acquired in axial plane with four contrasts: T2-weighted (AXT2), T1-weighted (AXT1), T1-weighted post-contrast (AXT1POST), and FLAIR (AXFLAIR). The data contain 4469/1378/558 2D volumes for the train/validation/test portions. Unlike the train/validation sets, the test set contains only the under-sampled k-spaces (and not fully sampled). As a common practice in the leaderboard [34, 11], we combined the train/validation sets to create a larger training set. Specifically, we created our split of 5647/200 volumes for train/validation portions.

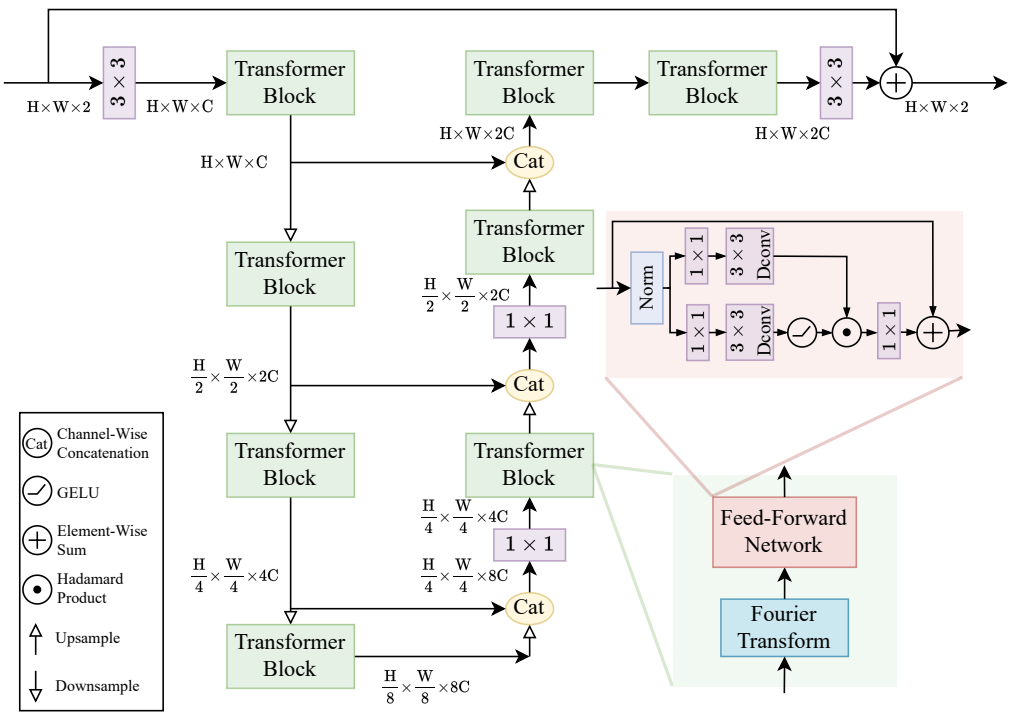

Figure 2: **Fourier Restormer is the backbone neural network of our approach.** Our variant of the Restormer [45] architecture, which is a U-Net style transformer. Our network, dubbed FRestormer, is based on parameter-free Fourier blocks instead of learnable attention modules which makes it amenable to high-resolution medical imaging. The input and output have 2 channels representing the real and imaginary parts of the image.

Comparison Methods: We compare our model to the top-5 publicly available models on the leaderboard, namely DIRCN [28], E2E-VarNet [34], XPDNet [31], HQS-Net [41], and U-Net [17, 46]. All of these models are unrolled networks except U-Net which is a CNN without any form of data consistency. For our model submission, we also used a self-ensemble technique based on horizontal flipping. Specifically, the forward pass of our model (at training and inference) is modified such that it computes the average of two output images; one for the input image, and one for the flipped input image (please see Sec. 3.3 for more details). We refer to this variant of our model as IR-FRestormerF which is a combination of IR-FRestormer and "Flip ensemble."

Metric: We evaluate reconstruction accuracy via the structural similarity index measure (SSIM) [40]. SSIM is widely used in the literature to evaluate the performance of image reconstruction algorithms and is the primary metric in the fastMRI challenge [19, 27]. However, as our model appears on the fastMRI leaderboard, PSNR and NRMSE are shown there as well.

3.2. Results

In the development stage of our model, we conducted numerous comparisons between our model and other state-of-the-art models and found that our model performs best on the validation set of the fastMRI dataset (which is publicly available). In this section, we present leaderboard results. These results are obtained for the test set of the fastMRI dataset whose ground-truth images are not publicly available[3]. The dataset organizers conduct this evaluation for every new model that is submitted to the leaderboard.

Leaderboard track: We target the 8×-accelerated multi-coil brain track of the fastMRI leaderboard. This track is more challenging than 4×-accelerated multi-coil brain since it deals with a higher acceleration factor. It is also more challenging than the 8×-accelerated multi-coil knee track since it has 4 imaging modalities (the knee dataset has 2), it has 5 times more volumes than the knee dataset, and also from the computational perspective, brain samples are larger in size.

[3]At the time of submission, the test set was released to the public, but our model was evaluated apriori on the leaderboard.

Model	SSIM	#params
DIRCN [28]	0.9455	45M
IR-FRestormerF (ours)	0.9427	2.5M
E2E-VarNet [34]	0.9426	30M
XPDNet [31]	0.9408	-[4]
HQS-Net [41]	0.9400	33M
U-Net [46]	0.9282	214M

Table 2: **Our IR-FRestormerF is among state-of-the-arts for 8x accelerated multi-coil brain MRI reconstruction.** Test SSIM and number of parameters of several of the best-performing models on the leaderboard. Our IR-FRestormer achieves state-of-the-art performance with significantly less number of parameters compared to its competitors on the leaderboard.

Leaderboard results: The leaderboard results shown in Table 2 demonstrate that our method achieves on-par performance with state-of-the-art models for 8×-accelerated multi-coil brain MRI reconstruction. This performance is achieved with fewer number of parameters compared to the other leaderboard models.

Aside from the quantitative analysis above, we also conducted a qualitative comparison by visualizing the reconstructions. Note that this qualitative comparison is naturally challenging in our work since the difference between top-performing methods is extremely small as also reflected in the quantitative metrics. Thus, it is difficult to highlight a major difference in the reconstructed images. Figure 3 shows reconstruction examples. Among the best-performing methods, we selected E2E-VarNet and our method, and as a baseline, we selected U-Net for visualization. As shown, E2E-VarNet and our method achieve on-par visual quality and both outperform the baseline U-Net. Furthermore, our model performs best in recovering fine details.

3.3. Ablation studies

For our model development, we conducted numerous ablations. The major elements of our method are FRestormer and iterative refinement (and a self-ensemble technique we used for our final submission which is described in the following).

Data split: For the purpose of these ablation studies, we created a sufficiently large subset of the train/validation portions of the multi-coil brain dataset since train and validation sets contain fully-sampled images. The subset we created is the same used in Section 2.2 which contains 500/170/140 volumes for train/validation/test portions.

[4]The number of model parameters for this method is not published by the authors.

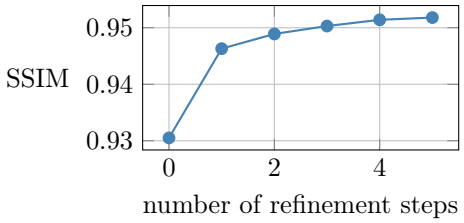

Figure 4: The effect of number of refinement steps on the performance of IR-FRestormer.

FRestormer architecture: As discussed in Section 2.2, we initially selected the Restormer architecture as for the choice of neural network in our framework. We proposed our variant FRestormer for the sake of memory efficiency. Table 3 shows that Restormer and FRestormer achieve similar accuracy on our test subset, but FRestormer consumes 2× less GPU memory for training. This is critical as the neural network in our iterative refinement framework should be highly memory efficient and its inefficiency results in having less number of refinement steps. FRestormer is also more efficient for storage purposes because of its small size.

Iterative Refinement with FRestormer: Another ablation, also shown in Table 3, compares IR-Restormer with IR-FRestormer. Not that in this experiment, IR-FRestormer has 3 refinement steps whereas IR-Restormer has 1 and both occupy the same 40 GB of GPU memory during training. Further, IR-FRestormer significantly outperforms IR-Restormer with $p < 0.01$ after performaing a paired Wilcoxon signed-rank test.

IR-FRestormer with flip ensemble: For our final submission, we also added flip ensemble to our model. For flip ensemble, during training, for each sample, we pass the input $\mathbf{x}$ and its horizontally-flipped version flip($\mathbf{x}$) and then take the average of the resulting outputs as the final output (f($\mathbf{x}$) + flip(f(flip($\mathbf{x}$))))/2. We then use the same forward pass during inference as well. Our flip ensemble is a variant of self-ensemble techniques [24, 25] which is very common in super-resolution challenges [3]. The last row in Table 3 shows the result of adding flip ensemble to IR-FRestormer. This self-ensemble technique adds improvement to the reconstruction accuracy and we used that in our final leaderboard submission as well.

For the ablations explained above, Fig. 3 shows reconstruction examples. An important remark is that IR-FRestormerF outperforms the rest of the models in terms of recovering fine details in the image.

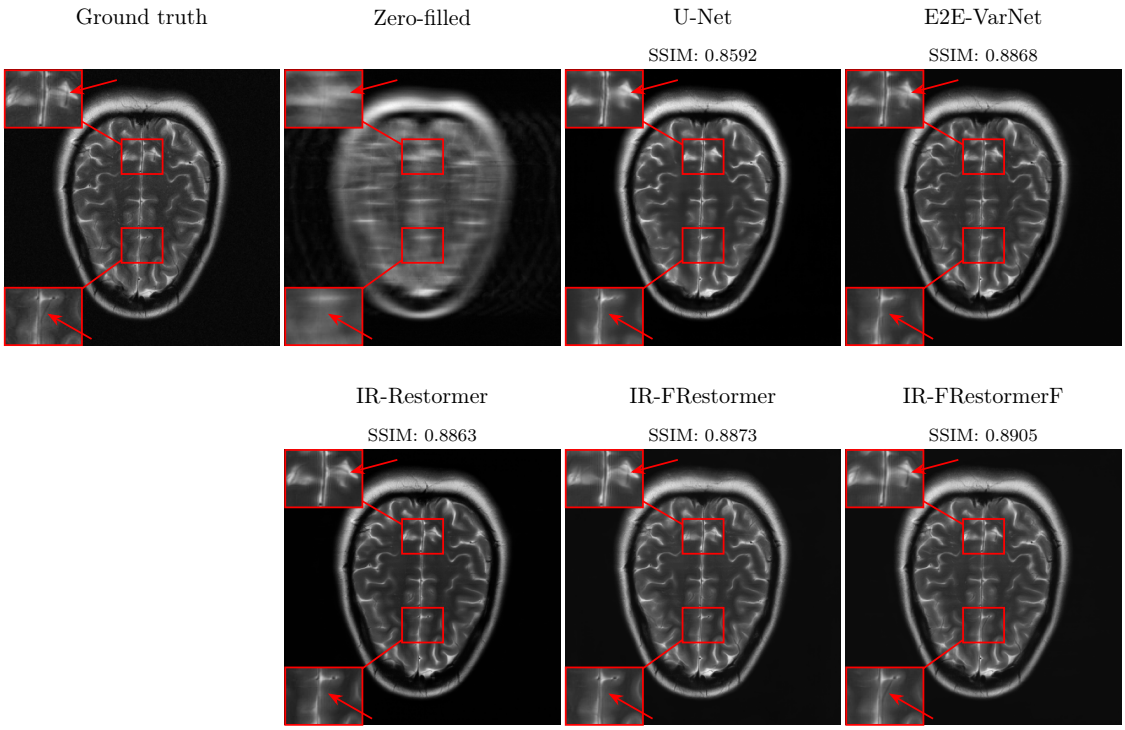

Figure 3: Reconstruction examples along with the zero-filled and ground-truth images for an AXT2 validation sample of multi-coil brain set (8×-acceleration). The first row contains U-Net and E2E-VarNet to which we compare our method and the second row contains our final model IR-FRestormerF along with our ablations IR-Restormer and IR-FRestormer. For this example, IR-FRestormerF performs best in recovering fine details in the image.

Model	SSIM	p-value	GPU Mem (GB)
Restormer	0.9319	-	12
FRestormer	0.9322	0.8	6
IR-Restormer	0.9491	<0.01	40
IR-FRestormer	0.9503	<0.01	40
IR-FRestormerF	**0.9532**	<0.01	40

Table 3: **IR-FRestormerF (IR-FRestormer + Flip-ensemble) performs best w.r.t the SSIM score.** Our ablations show that iterative refinement, Fourier-based attention, and flip-ensemble all result in performance improvement. Each p-value is the result of paired Wilcoxon signed-rank test between the setup of the row that contains the p-value and its previous row. Note that with our FRestormer, IR-FRestormer can have 3 refinement steps compared to IR-Restormer which has only one refinement step with the same training GPU memory cost on an A40 GPU.

Number of refinement steps: Our final ablation investigates the role of number of refinement steps in the performance. For this purpose, we varied the number of refinement steps from 0 to 5 and as shown in Fig. 4, the performance monotonically increases and saturates after 3 refinement steps. Thus, there is a trade-off between performance and memory efficiency as more number of refinement steps means more memory consumption.

4. Conclusion

Accelerated MRI reconstruction is an important topic in medical imaging. For this task, the most challenging track of the popular fastMRI leaderboard, 8×-accelerated multi-coil brain MRI reconstruction, has been advanced only once over the past two years.

In this work, we proposed a new state-of-the-art model based on combining a new efficient transformer-based architecture with iterative refinement. Our model achieved on-par performance with other state-of-the-art methods and recovered fine details with better quality.

Our work shows that iterative refinement can be a promising framework for accelerated MRI reconstruction, and transformer-based networks can be used for this task after addressing their memory consumption issues with Fourier components. This opens up new possibilities for more future direction on these two methods in accelerated MRI reconstruction.

References

[1] H. K. Aggarwal, M. P. Mani, and M. Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2018. 2, 4

[2] S. Arora, V. Roeloffs, and M. Lustig. Untrained modified deep decoder for joint denoising parallel imaging reconstruction. In *International Society for Magnetic Resonance in Medicine*, 2020. 2

[3] G. Bhat, M. Danelljan, R. Timofte, Y. Cao, et al. NTIRE 2022 burst super-resolution challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1041–1061, 2022. 7

[4] A. Björck. Iterative refinement of linear least squares solutions. In *BIT Numerical Mathematics*, volume 7, pages 257–278, 1967. 4

[5] K. T. Block, M. Uecker, and J. Frahm. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. In *International Society for Magnetic Resonance in Medicine*, pages 1086–1098, 2007. 1

[6] J. Cheng, H. Wang, L. Ying, and D. Liang. Model learning: Primal dual networks for fast MR imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 21–29, 2019. 4

[7] M. Zalbagi Darestani and R. Heckel. Accelerated MRI with un-trained neural networks. In *IEEE Transactions on Computational Imaging*, volume 7, pages 724–733, 2021. 2

[8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2020. 3

[9] T. Eo, Y. Jun, T. Kim, J. Jang, HJ. Lee, and D. Hwang. KIKI-Net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. volume 80, pages 2188–2201, 2018. 3

[10] K. Epperson, AM. Sawyer, M. Lustig, M. Alley, and M. Uecker. Creation of fully sampled MR data repository for compressed sensing of the knee. In *Annual Meeting for Section for Magnetic Resonance Technologists*, 2013. 1

[11] Z. Fabian and M. Soltanolkotabi. HUMUS-Net: Hybrid unrolled multi-scale network architecture for accelerated MRI reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 5

[12] CM. Feng, Y. Yan, G. Chen, H. Fu, Y. Xu, and L. Shao. Accelerated multi-modal MR imaging with transformers. In *arXiv preprint:2106.14248 [cs, eess]*, 2021. 3

[13] CM. Feng, Y. Yan, H. Fu, L. Chen, and Y. Xu. Task transformer network for joint MRI reconstruction and super-resolution. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 307–317, 2021. 3

[14] M. Genzel, I. Gühring, J. Macdonald, and M. März. Near-exact recovery for tomographic inverse problems via deep learning. In *International Conference on Machine Learning (ICML)*, pages 7368–7381, 2022. 4

[15] R. Heckel and P. Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2019. 2

[16] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir. Robust compressed sensing MRI with deep generative priors. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021. 3

[17] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. In *IEEE Transactions on Image Processing*, pages 4509–4522, 2017. 3, 6

[18] V. A. Kelkar and M. Anastasio. Prior image-constrained reconstruction using style-based generative models. In *International Conference on Machine Learning (ICML)*, pages 5367–5377, 2021. 3

[19] F. Knoll, T. Murrell, A. Sriram, N. Yakubova, J. Zbontar, M. Rabbat, A. Defazio, M. J. Muckley, D. K. Sodickson, C. L. Zitnick, et al. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. In *Magnetic Resonance in Medicine*, 2020. 2, 6

[20] Y. Korkmaz, S. UH. Dar, M. Yurt, M. Özbey, and T. Cukur. Unsupervised MRI reconstruction via zero-shot learned adversarial transformers. In *IEEE Transactions on Medical Imaging*, 2022. 3

[21] Y. Korkmaz, M. Yurt, S. UH. Dar, M. Özbey, and T. Cukur. Deep MRI reconstruction with generative vision transformers. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 54–64, 2021. 3

[22] J. Lee-Thorp, J. Ainslie, I. Eckstein, and S. Ontanon. Fnet: Mixing tokens with fourier transforms. In *arXiv preprint:2105.03824 [cs.CL]*, 2021. 5

[23] K. Lin and R. Heckel. Vision transformers enable fast and robust accelerated MRI. In *Medical Imaging with Deep Learning (MIDL)*, 2021. 3

[24] J. Liu, CH. Wu, Y. Wang, Q. Xu, et al. Learning raw image denoising with bayer pattern unification and bayer preserving augmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019. 7

[25] Z. Luo, L. Yu, X. Mo, Y. Li, L. Jia, H. Fan, J. Sun, and S. Liu. EBSR: Feature enhanced burst super-resolution with deformable alignment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 471–478, 2021. 7

[26] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. In *International Society for Magnetic Resonance in Medicine*, pages 1182–1195, 2007. 1

[27] M. J. Muckley, B. Riemenschneider, A. Radmanesh, S. Kim, G. Jeong, J. Ko, Y. Jun, H. Shin, D. Hwang, M. Mostapha, et al. State-of-the-art machine learning MRI reconstruction in 2020: Results of the second fastMRI challenge. In *IEEE Transactions on Medical Imaging*, 2021. 2, 6

[28] J. A. Ottesen, M. Caan, I. R. Groote, and A. Bjørnerud. A densely interconnected network for deep learning accelerated MRI. In *Magnetic Resonance Materials in Physics, Biology and Medicine*, pages 1–13, 2022. 6, 7

[29] N. Parikh, S. Boyd, et al. Proximal algorithms. In *Foundations and trends® in Optimization*, volume 1, pages 127–239, 2014. 4

[30] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. In *International Society for Magnetic Resonance in Medicine*, volume 42, pages 952–962, 1999. 5

[31] Z. Ramzi, P. Ciuciu, and JL. Starck. XPDNet for MRI reconstruction: An application to the 2020 fastMRI challenge. In *arXiv preprint:2010.07290 [eess.IV]*, 2020. 6, 7

[32] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015. 3, 5

[33] Y. Song, L. Shen, L. Xing, and S. Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations (ICLR)*, 2022. 3

[34] 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 *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 64–73, 2020. 2, 3, 5, 6, 7

[35] J. Sun, H. Li, Z. Xu, et al. Deep ADMM-Net for compressive sensing MRI. *Advances in Neural Information Processing Systems (NeurIPS)*, 29, 2016. 4

[36] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. In *Magnetic Resonance in Medicine*, pages 990–1001, 2014. 5

[37] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9446–9454, 2018. 2

[38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017. 3

[39] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. Accelerating magnetic resonance imaging via deep learning. In *IEEE international Symposium on Biomedical Imaging (ISBI)*, pages 514–517, 2016. 1, 2

[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. In *IEEE Transactions on Image Processing*, pages 600–612, 2004. 6

[41] B. Xin, T. S. Phan, L. Axel, and D. N. Metaxas. Learned half-quadratic splitting network for MR image reconstruction. In *Medical Imaging with Deep Learning (MIDL)*, 2021. 6, 7

[42] B. Yaman, SAH. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. volume 84, pages 3172–3191, 2020. 2

[43] W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, and S. Yan. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10819–10829, 2022. 5

[44] M. Yurt, M. Özbey, S. Dar, B. Tinaz, K. Oguz, and T. Çukur. Progressively volumetrized deep generative models for data-efficient contextual learning of MR image recovery. In *Medical Image Analysis*, volume 78, 2022. 3

[45] SW. Zamir, A. Arora, S. Khan, M. Hayat, FS. Khan, and MH. Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5728–5739, 2022. 5, 6

[46] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, et al. fastMRI: An open dataset and benchmarks for accelerated MRI. In *Radiology: Artificial Intelligence*, 2020. 1, 2, 5, 6, 7

[47] J. Zhang and B. Ghanem. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1828–1837, 2018. 4