World Academy of Science, Engineering and Technology
International Journal of Mechanical and Mechatronics Engineering
Vol:17, No:8, 2023

End-to-End Pyramid Based Method for MRI Reconstruction

Omer Cahana, Maya Herman, Ofer Levi

Abstract—Magnetic Resonance Imaging (MRI) is a lengthy medical scan that stems from a long acquisition time. Its length is mainly due to the traditional sampling theorem, which defines a lower boundary for sampling. However, it is still possible to accelerate the scan by using a different approach such as Compress Sensing (CS) or Parallel Imaging (PI). These two complementary methods can be combined to achieve a faster scan with high-fidelity imaging. To achieve that, two conditions must be satisfied: i) the signal must be sparse under a known transform domain, and ii) the sampling method must be incoherent. In addition, a nonlinear reconstruction algorithm must be applied to recover the signal. While the rapid advances in Deep Learning (DL) have had tremendous successes in various computer vision tasks, the field of MRI reconstruction is still in its early stages. In this paper, we present an end-to-end method for MRI reconstruction from k-space to image. Our method contains two parts. The first is sensitivity map estimation (SME), which is a small yet effective network that can easily be extended to a variable number of coils. The second is reconstruction, which is a top-down architecture with lateral connections developed for building high-level refinement at all scales. Our method holds the state-of-art fastMRI benchmark, which is the largest, most diverse benchmark for MRI reconstruction.

Keywords—Accelerate MRI scans, image reconstruction, pyramid network, deep learning.

I. Introduction

MAGNETIC resonance imaging (MRI) is a type of non-invasive scan used to visualize internal organs in the body without exposing patients to ionizing radiation. This method uses a combination of radio radiation and a strong magnetic field to influence the spin of the protons in the hydrogen atoms found in the body's water molecules. This change in the spin properties depends on where the tissue is located. By sending a pulse of radio waves, the MRI creates a sharp movement in the proton spin, causing the proton to create a magnetic resonance that the machine's input coils can measure. Using a Fourier transform (FT), which is the most common transformation in MRI machines, the signal is separated by frequency from the phase component as a function of time. This plane is also known as the k-space. Two methods have been developed to accelerate the acquisition process: Parallel imaging (PI) [1] and compress sensing (CS) [2], [3]. PI is a technique to reduce time-consuming phase-encoding steps that use the knowledge of the receiver coil placement and its sensitivity to create a special localization signal. PI produces a complex matrix for each input coil, which is multiplied by the position-dependent coil sensitivity map. Those maps are usually generated from the auto-calibration signal (ACS) that corresponds to the low spatial frequencies. PI is divided into two main approaches. The first category includes approaches that take place in the image domain, such as sensitivity encoding (SENSE) [4], in which coil sensitivities are used to sort out the signals after Fourier transform. The second category includes approaches that take place in the k-space domain, such as generalized auto-calibrating partial parallel acquisition (GRAPPA) [5], in which missing harmonic data are corrected before applying the Fourier transform. In practice, those methods are limited to a factor of three due to noise being significantly enhanced in this process. CS is also based on the premise that an image can be reconstructed from an incompletely filled k-space. However, despite PI, CS does not use complimentary information. CS suggests that signals can be recovered with fewer samples than the traditional sampling theorem. For it to be applied, the signabasis pursuitl must be sparse in a known transform domain. Additionally, the sampling must be incoherent to eliminate artifacts such as the aliasing artifact. While MR images are rarely sparse, one important observation has been applied to CS-MRI: signal redundancy is closely related to sparsity [6]. This is true because redundant signals can be converted to sparse signals using known transforms. In this paper, we present a method for MRI reconstruction. Our method utilizes PI by using the SME network, whereas the reconstruction part utilized the CS-MRI. In summary, our contributions are as follows: i) a method for sensitivity map estimation, ii) a method for reconstructing a network with two-stage lateral connection, and iii) state-of-the-art results on the fastMRI benchmark.

Prof. Ofer Levi, Dr. Maya Herman, and Omer Cahana are with Department of Computer Science, The Open University, Israel (e-mail: oferle@openu.ac.il, maya@openu.ac.il, omer4436@gmail.com).

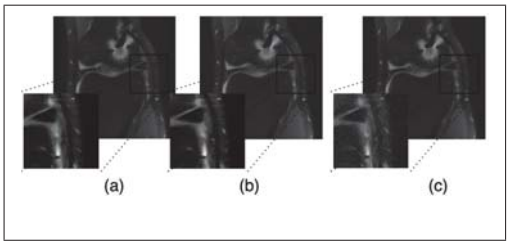

Fig. 1 Qualitative results on knee sample: (a) our method, (b) VarNet, (c) target scan

II. Related Work

Deep learning is a promising technique for a wide variety of ML problems. As a result of such success, deep learning has been applied to computational MRI problems, showing

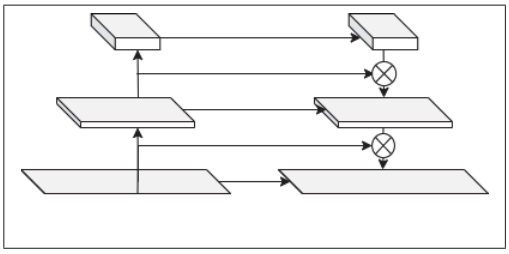

Fig. 2 The proposed method as a pyramid-based approach that involves down-up stages

promise as a means to speed up MR reconstruction. Putzky et al. suggest using the invertible variant of Recurrent Inference Machines (RIM) [12], an iterative map combined with the current reconstruction, a hidden memory state, and the gradient of the likelihood term; it encodes information about the known generative process and measures how well it reproduces those measurements. This kind of work, which can be viewed as an inverse problem, holds an unrealistic assumption that the forward pass in multi-coil MRI is completely known (it is determined by the sensitivity maps). Jeong et al. suggest multi-domain data processing and data standardization. They performed the process of standardization at the multichannel coil level, which was processed using GRAPPA [5] on the oversampled area in the image domain. ESPIRiT [13] sensitivity maps were used to reorganize preprocessed data into a channel combination and residual images, which created a consistent data format. They used the same convolutional layers for multi-domain data processing that shares its features between the domain, the k-space, and the image. Ramzi et al. suggest the network XPDNet [14], which is an end-to-end unrolled cross-domain network based on the Chambolle-Pock algorithm. They used the last five unrolled iterations to ensure learning a complex non-linear acceleration scheme. Hammernik et al. proposed a trainable formula that generalized CS as a Variational Network (VN) [15]. VN initially developed into classic image processing tasks as a trainable reaction-diffusion model. The VN model is a fully learned model that is based on the incremental proximal gradient method. Later, Sriram et al. extended Hammernik's work, proposing a VN model that learns completely end-to-end [16]. Additionally, they proposed neural network architectures that determine the optimal intermediate representation for better reconstruction.

III. Method

We present a pyramid-based method with a lateral connection that leverages the information from low to high levels, capturing fine details to improve image reconstruction. Our method is composed of two main networks: the sensitivity map estimation (SME) and the reconstruction network. While an SME network is a small yet efficient network, the reconstruction network consists of three main parts: the aggregate block, the de-aggregate block, and the reconstruction block. Both networks receive the same input and the same k-space. Additionally, they are connected after the reconstruction block throughout the de-aggregation to the output. We call each presented iteration a stage. The upsampling part of the pyramid has a lateral connection with the down-sampling part to refine the k-space based on low-level information that learns from the parallel stage. Finally, the overall method is presented in Fig. 2 and defined in (1). Each component is described below:

$$I^{i+1} = DeAgg \circ (Recon(Agg(I^i)) * SME(I^i)) \quad (1)$$

A. Sensativity Map Estimation

In the last few years, few works suggest replacing the ESPIRiT algorithm, which was the most popular algorithm for coil sensitivity estimation in CNN-based algorithms. One of them was presented in a VarNet paper, in which they used a simple U-net model to learn the coil sensitivity. We leverage their idea with a few changes. First, we suggest using a much smaller network, since there is no need for rich features and a heavy model to estimate sensitivity. Second, we leverage the advantage of dilated convolution to extract localization information, further reducing the number of features. For the sake of paper completeness, we present the SME method formula:

$$dSS \circ CNN \circ \mathcal{F}^{-1} \circ M_{center} \quad (2)$$

where M center corresponds to ACS lines and dSS[vanet] is a normalization operator.

B. Reconstruction Block

The reconstruction block is defined as CNN. The network that we use is a modification of U-net where the latent space is a concatenation of two feature maps, the one from the last downsampling layer and the one from the previous network. This cross-stage lateral information helps to extend the information flow throughout the method and refined the fine details properly.

C. Aggregate Block

The aggregate block contains two main functions. The first is an inverse Fourier transform that maps the input and the k-space in the frequency domain to the spatial space. The second is the aggregation of the input, which is represented with multiple coils for each image. This process can be done using two methods: root sum of square (RSS), or root mean square (RMS). The ablation study contains a comparison of these methods.

D. De-Aggregate Block

The de-aggregate block is conceptually the inverse of the aggregate block. First, we use a Fourier transform, then multiply the sensitivity map from the SME block by the output of the reconstruction block.

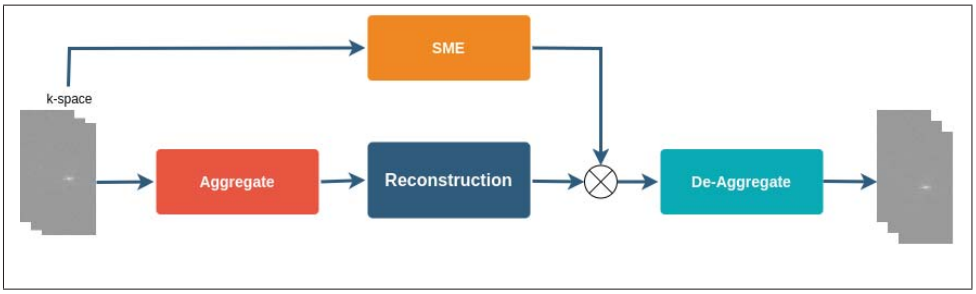

Fig. 3 One stage in out method from K-space to image reconstruction

TABLE I
QUANTITATIVE RESULTS ON THE FASTMRI TEST SET

Scan	Model	Acceleration Factor = 4			Acceleration Factor = 8		
		SSIM	NMSE	PSNR	SSIM	NMSE	PSNR
Brain	Unet	0.946	0.007	38	0.922	0.014	35
	XPD-net	0.959	0.003	41	0.942	0.007	38
	VarNet	0.959	0.004	41	0.943	0.007	38
	AIRS-net	0.964	0.003	42	0.952	0.005	40
	Our Method	**0.968**	0.003	**43**	**0.954**	0.005	**41**
Knee	Unet	0.91	0.007	34	0.864	0.013	35
	i-RIM	0.928	0.005	40	0.888	0.009	37
	SigmaNet	0.928	0.005	40	0.888	0.009	37
	AIRS Medical	0.929	0.005	40	0.888	0.009	37
	SubtleMR	0.929	0.005	40	0.888	0.009	37
	VarNet	0.930	0.005	40	0.890	0.009	37
	Our Method	**0.932**	0.005	**41**	**0.895**	0.009	**38**

IV. EXPERIMENTS

A. MRI Scan Sequences

The signal obtained from the magnetic resonance process is produced by the protons located within the nuclei of hydrogen atoms through their procession. This signal is made possible by the magnetic field moving in phases separately from the external magnetic field. Moreover, with changes in repetition time (TR) and echo time (TE), we can determine the factor which will most influence the signal. Then we can get some images in which both the tissues and the pathologies are differentiated based on features that characterize their components. Contrast is required to distinguish normal anatomy from pathology. Contrast is improved when two adjacent areas have high and low signal intensities, respectively. There are many different MRI sequences (more than 100), all of which attempt to optimize tissue contrast. Each MR imaging consists of a T1 component and a T2 component. It is possible to switch off most of either component, creating either a T1-weighted image, a T2-weighted image or T1- POST or FLAIR image. Proton density (PD) weighted images are a special type of MRI. PD-weighted images visualize the number of protons per volume; tissues with fewer protons have lower signal intensity, while tissues with many protons have higher signal intensities. PD-weighted images are used to evaluate meniscal tears in the knee, among other uses.

B. Dataset

In this study, we evaluate our model and sampling method using the Fast MRI dataset, which contains brain and knee MR imaging. A knee scan is acquired with two pulse sequences: proton density (PD) and proton density with fat suppression (PDFS) weighted images, which increase contrast. The knee dataset contains 1,539 scans and the brain dataset contains 6,970 scans; both datasets are acquired from both the 3T and 1.5T systems. There are two tracks in this dataset: multi-coil, which contains 15 channels, and single coil, which contains one coil. Brain scans are acquired with four pulse sequences, such as T1, T2, T1 POST, and FLAIR-weighted images. There are a number of changes in those scans—both in pathology and in the image—which contain different edges, contrasts, and brightnesses.

C. Evaluation Metrics

We used three metrics that are commonly used in the literature for image reconstruction to evaluate our method: structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), and normalized mean square error (NMSE).

SSIM—attempts to quantify the perceived image quality, which is define as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy}^2 + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

where μ_x, μ_y are the average pixel intensities in x and y, respectively, σ_x, σ_y are their variances, σ_{xy} is the covariance between x and y, and c_1, c_2 are two variables to stabilize the division. For a sanity check, we use a VarNet validation with window size of 7x7, k_1=0.001, k_1=0.003 and $L = max(v)$, where v stands for the maximum value of the target volume.

TABLE II
ABLATION STUDY ON THE IMPROVEMENT THAT WE SUGGESTED

base model	lateral output	lateral latent	SSIM	NMSE	PSNR
X			0.891	0.009	37
X	X		0.892	0.010	37
X		X	0.894	0.010	38
X	X	X	0.895	0.010	38

TABLE III
COMPARISON VARNET VS OUR METHOD ON FLAIR SCANS

Model	SSIM	NMSE	PSNR
VarNet	0.932	0.007	39
Our	0.956	0.004	42

NMSE—a pixel-wise metric that computes the error between a reconstructed image, x, to target image, y, which is defined as follows:

$$NMSE(x,y) = \frac{||x-y||_2^2}{||y||_2^2} \tag{4}$$

PSNR—widely used in Compress Sensing (CS) applications to evaluate the quality of reconstructions, which are defined as follows:

$$PSNR(x,y) = 10\frac{y^2}{MSE(x,y)} \tag{5}$$

D. Implementation Details

We designed our method with six stages, where the loss function is a linear combination of SSIM with l1-loss. We chose alpha to be 0.3, empirically. We trained with AdamW as an optimizer with a learning rate of 0.1 with scheduler. For the activation function, we used leaky-ReLU without dropout. Additionally, we used instance normalization since we feed only one sample every iteration. The training was performed on two NVIDIA A100 GPUs with 40GB each.

$$L = \alpha * l_1 + (1-\alpha) * (1 - SSIM) \tag{6}$$

$$\mathcal{L} = \alpha \cdot l_1 + (1-\alpha) \cdot (1 - SSIM) \tag{7}$$

V. RESULTS

Table I shows the results of our experiments on the fastMRI benchmark. In the knee dataset, we use random sampling. However, in the brain dataset, we use equidistant sampling as suggested in fastMRI paper. We evaluate on two-factor acceleration: 4x and 8x. On the brain dataset, where most of the data are centralized, we achieved better performance using SSIM. However, we found that results using the NMSE and PSNR were equivalent to those using AIRS-net. In the knee dataset, our method performed better on all the evaluation metrics by a large margin. In Table II, we compare different parts of our method and how well they contribute to the overall method. Its most significant part is the concatenation of two latent spaces. In Table III, we examine the influence of variation and limited data (FLAIR scans) on the overall result. We infer that the performance decreases due to a lack of generalization in the FLAIR scan since it is smaller and different than other scans.

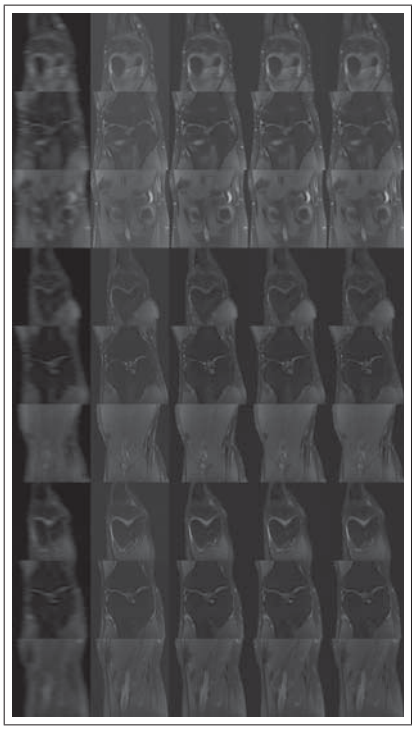

Fig. 4 Qualitative comparison left to right: input, unet, vernet, our method, target on knee dataset with 4x acceleration

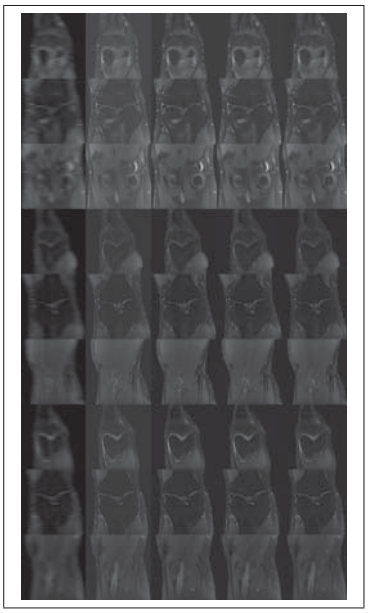

Fig. 5 Qualitative comparison left to right: input, unet, vernet, our method, target on knee dataset with 8x acceleration

VI. DISCUSSION AND LIMITATIONS

Supervised learning generalizes information from a certain distribution. However, the scans may change considerably

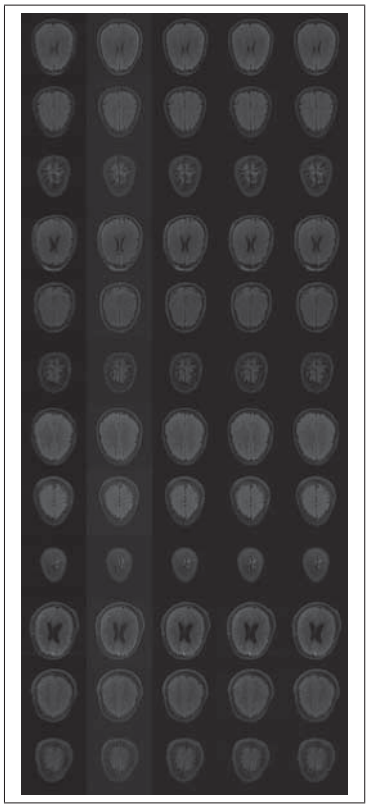

Fig. 6 Qualitative comparison left to right: input, unet, vernet, our, target on brain dataset with 4x acceleration

following a change in their parameters such as a change in TE or TR, etc. To properly reconstruct a scan, we must train the models from scratch for each type of machine and for a set of parameters that relatively fixed. A second point is the reliability of the results; the methods we have presented have been measured in numerical values which do not necessarily correspond to the quality of the reconstruction as determined by pathologists. To bridge this gap, the research community must also collect pathological values for diagnosing the results. Finally, to provide a truly comprehensive, applicable, and reliable solution, we need to publish more benchmarks with raw information, additional types of scans, various parameters, and various degrees of freedom.

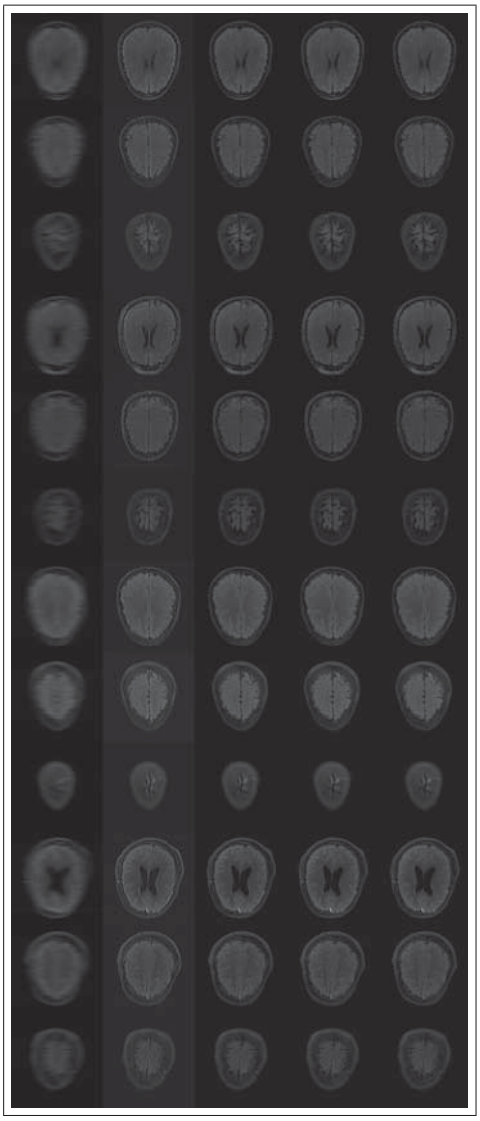

Fig. 7 Qualitative comparison left to right: input, unet, vernet, our, target on brain dataset with 8x acceleration

VII. Conclusions

In this paper, we have introduced a method for reconstructing MR scans. Our method consists of two main components: the reconstruction network, which iteratively learns the reconstruction, and SME network, which learns the sensitivity map. Moreover, we introduce the latent space lateral connection to learn the refinement details. Finally, we show an ablation study on our improvements and compared our results to the last state-of-the-art model in fastMRI [17].

References

[1] A. C.-Y. Yang, M. Kretzler, S. Sudarski, V. Gulani, and N. Seiberlich, "Sparse reconstruction techniques in MRI: methods, applications, and challenges to clinical adoption," *Investigative radiology*, vol. 51, no. 6, pp. 349, 2016, NIH Public Access.

[2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008, IEEE.

[3] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*, Cambridge university press, 2012.

[4] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999, Wiley Online Library.

[5] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002, Wiley Online Library.

[6] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 59, no. 2, pp. 365–373, 2008, Wiley Online Library.

[7] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Compressed-sensing MRI with random encoding," *IEEE transactions on Medical Imaging*, vol. 30, no. 4, pp. 893–903, 2010, IEEE.

[8] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss, "Variable density sampling with continuous trajectories. Application to MRI," *arXiv preprint arXiv:1311.6039*, 2013.

[9] D. M. Spielman, J. M. Pauly, and C. H. Meyer, "Magnetic resonance fluoroscopy using spirals with variable sampling densities," *Magnetic resonance in medicine*, vol. 34, no. 3, pp. 388–394, 1995, Wiley Online Library.

[10] J. P. Haldar and D. Kim, "OEDIPUS: An experiment design framework for sparsity-constrained MRI," *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1545–1558, 2019, IEEE.

[11] C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-learning-based optimization of the under-sampling pattern in MRI," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1139–1152, 2020, IEEE.

[12] P. Putzky, D. Karkalousos, J. Teuwen, N. Miriakov, B. Bakker, M. Caan, and M. Welling, "i-RIM applied to the fastMRI challenge," *arXiv preprint arXiv:1910.08952*, 2019.

[13] 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," *Magnetic resonance in medicine*, vol. 71, no. 3, pp. 990–1001, 2014, Wiley Online Library.

[14] Z. Ramzi, P. Ciuciu, and J.-L. Starck, "XPDNet for MRI Reconstruction: an application to the 2020 fastMRI challenge," *arXiv preprint arXiv:2010.07290*, 2020.

[15] 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," *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018, Wiley Online Library.

[16] 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*, pp. 64–73, 2020, Springer.

[17] J. Zbontar et al., "fastMRI: An open dataset and benchmarks for accelerated MRI," *arXiv preprint arXiv:1811.08839*, 2018.