

Learning Task-Specific Strategies for Accelerated MRI

Zihui Wu¹, Graduate Student Member, IEEE, Tianwei Yin², Yu Sun, Member, IEEE, Robert Frost³, Andre van der Kouwe, Senior Member, IEEE, Adrian V. Dalca⁴, and Katherine L. Bouman⁵

Abstract—Compressed sensing magnetic resonance imaging (CS-MRI) seeks to recover visual information from subsampled measurements for diagnostic tasks. Traditional CS-MRI methods often separately address measurement subsampling, image reconstruction, and task prediction, resulting in a suboptimal end-to-end performance. In this work, we propose TACKLE as a unified co-design framework for jointly optimizing subsampling, reconstruction, and prediction strategies for the performance on downstream tasks. The naïve approach of simply appending a task prediction module and training with a task-specific loss leads to suboptimal downstream performance. Instead, we develop a training procedure where a backbone architecture is first trained for a generic pre-training task (image reconstruction in our case), and then fine-tuned for different downstream tasks with a prediction head. Experimental results on multiple public MRI datasets show that TACKLE achieves an improved performance on various tasks over traditional CS-MRI methods. We also demonstrate that TACKLE is robust to distribution shifts by showing that it generalizes to a new dataset we experimentally collected using different acquisition setups from the training data. Without additional fine-tuning,

TACKLE leads to both numerical and visual improvements compared to existing baselines. We have further implemented a learned $4\times$ -accelerated sequence on a Siemens 3T MRI Skyra scanner. Compared to the fully-sampling scan that takes 335 seconds, our optimized sequence only takes 84 seconds, achieving a four-fold time reduction as desired, while maintaining high performance.

Index Terms—Compressed sensing MRI, deep learning, end-to-end training, task-specific imaging.

I. INTRODUCTION

COMPRESSED sensing magnetic resonance imaging (CS-MRI) is a popular accelerated MRI technology [1]. Commonly, CS-MRI is formulated as an imaging inverse problem where the goal is to recover a high-quality image from its subsampled measurements. Traditional CS-MRI techniques include solving a regularized optimization problem [2], [3], [4], [5] or training deep learning (DL) models [6], [7], [8] that recover an image from a pre-determined set of measurements. Recently, a new group of DL-based methods, known as *co-design*, has been proposed to jointly optimize the choice of measurements and a reconstruction module, leading to better reconstruction performance than the traditional CS-MRI methods [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23].

In the existing co-design literature, task prediction is often viewed as a post-processing step decoupled from image reconstruction. All the aforementioned methods focus on image reconstruction and rely on standard image similarity metrics such as mean square error (MSE) or peak signal-to-noise ratio (PSNR) as a proxy for performance on a downstream task. Such a reconstruction-oriented formulation lacks a direct connection with the downstream tasks that reflect actual clinical needs [24]. We are thus motivated to ask: *can one improve the accuracy of downstream task prediction by optimizing the entire CS-MRI pipeline in an end-to-end fashion?*

With end-to-end co-design methods, it seems like we are only one step away from incorporating downstream tasks as part of the optimization. Namely, one can simply append a task prediction module and add a task-specific loss. However, as shown in Fig. 1 and Table II, this approach leads to a suboptimal performance on the task prediction and is sometimes even worse than the traditional approach of separate reconstruction and task prediction. These results indicate that it remains a challenge on how to robustly learn task-specific strategies for CS-MRI.

Manuscript received 6 November 2023; revised 7 March 2024 and 18 May 2024; accepted 21 May 2024. Date of publication 1 July 2024; date of current version 18 July 2024. The work of Zihui Wu was supported in part by Kortschak Fellowship, in part by Amazon AI4Science Fellowship, and in part by Amazon AI4Science Partnership Discovery Grant. This work was supported in part by NSF under Grant 2048237, in part by NIH under Grant 5R01AG064027, Grant 5R01AG070988, Grant R21EB029641, Grant R01HD099846, and Grant R01HD085813, in part by Heritage Medical Research Fellowship, in part by S2I Clinard Innovation Award, and in part by Rockley Photonics. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alejandro F. Frangi. (Corresponding author: Zihui Wu.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Mass General Brigham Institutional Review Board under IRB Protocol No. 2012P002376, and performed in line with the Department of Health and Human Services (DHHS) 45 CFR Parts 46 and 164.

Zihui Wu, Yu Sun, and Katherine L. Bouman are with the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91105 USA (e-mail: zwu2@caltech.edu; sunyu@caltech.edu; klbouman@caltech.edu).

Tianwei Yin is with the Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: tianwei@mit.edu).

Robert Frost and Andre van der Kouwe are with the Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, MGH, Harvard Medical School, Charlestown, MA 02129 USA (e-mail: rfrost@mgh.harvard.edu; avanderkouwe@mgh.harvard.edu).

Adrian V. Dalca is with the Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, and also with the Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, MGH, Harvard Medical School, Charlestown, MA 02129 USA (e-mail: adalca@mit.edu).

Our code is available at <https://github.com/zihuiwu/TACKLE>.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TCI.2024.3410521>, provided by the authors.

Digital Object Identifier 10.1109/TCI.2024.3410521

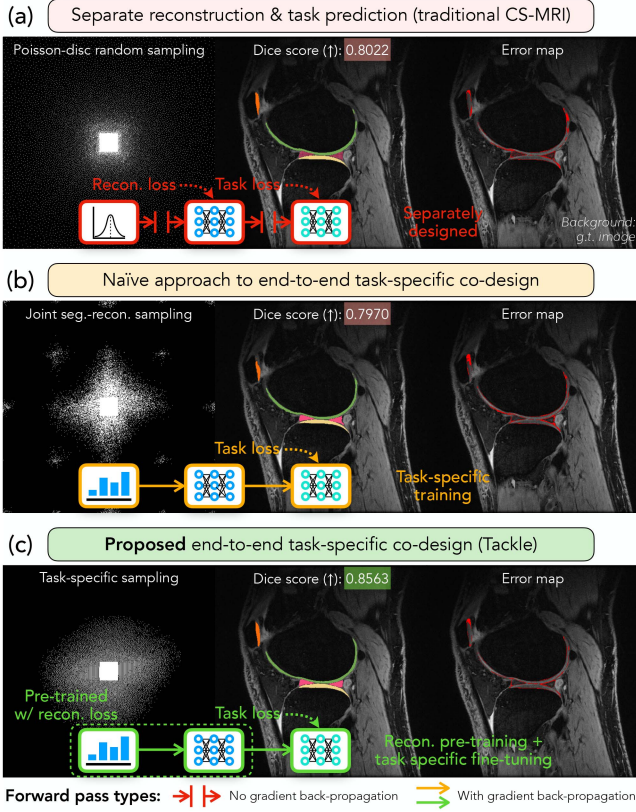


Fig. 1. Comparison between (a) traditional CS-MRI, (b) a naïve approach to task-specific CS-MRI, and (c) the proposed TACKLE framework. Compared with panel (a) that separately deals with reconstruction and task prediction, panel (b) is a simple extension of co-design methods for solving downstream tasks by adding a learnable mapping from measurements to task predictions. However, this naïve approach leads to a suboptimal performance and can even lead to a worse task prediction accuracy, as shown in the example above. On the other hand, we introduce TACKLE for effectively learning task-specific CS-MRI strategies. TACKLE is first pre-trained for generic reconstruction, and then all three modules are fine-tuned for a more specific downstream task. We find that this training schedule allows TACKLE to robustly learn generalizable task-specific strategies. In the above knee segmentation example, all three approaches are trained with the same architectures for the reconstructor (second module) and predictor (third module). Nevertheless, TACKLE significantly outperforms the two baseline approaches.

In this paper, we address this challenge by proposing a unified framework, **task-specific codesign of k -space subsampling and prediction** (TACKLE), for designing task-specific CS-MRI systems. Different from existing works that focus on specific tasks, TACKLE is a general framework that accommodates different downstream tasks. To do so, we design a two-step training strategy that mimics the training of modern language and vision models. TACKLE is first trained for a generic task of image reconstruction, and then fine-tuned for specific downstream tasks. We find that this approach can effectively learn generalizable task-specific strategies that lead to significant and consistent improvements, with an example shown in Fig. 1(c). Besides the standard task of reconstructing the full FOV (which we call full-FOV reconstruction hereafter), we demonstrate TACKLE on three other tasks covering both pixel-level and image-level imaging problems: region-of-interest (ROI) oriented reconstruction,

tissue segmentation, and pathology classification. Our experimental results show that end-to-end optimization for task prediction sometimes circumvents the typical reconstruction in terms of point-wise accuracy, but leads to improved accuracy on the task of interest by effectively extracting key visual information for task prediction.

The main contributions of this work are as follows:

- We provide a general framework (TACKLE) that learns specific strategies for a variety of CS-MRI tasks. TACKLE optimizes the entire CS-MRI pipeline, from measurement acquisition to label prediction, in an end-to-end fashion *directly* for a user-defined task.
- We validate TACKLE on multiple MRI datasets, covering different body parts, scanning sequences, and hardware setups. Experimental results show that TACKLE outperforms the reconstruction-oriented baseline methods on *all* considered settings. We evaluate the proposed end-to-end architecture and training procedure through ablation studies. Our results offer guidance for designing effective task-specific CS-MRI systems in the future.
- We show the generalization of TACKLE to out-of-distribution data by deploying it to a dataset we experimentally acquired using a different acquisition sequence from that of the training data. We further implement a learned $4\times$ -accelerated sequence on a Siemens 3T MRI Skyra scanner. The sequence shortens the scan time from 335 seconds to 84 seconds, a four-fold time reduction as desired, while maintaining high performance. These experiments highlight the real-world practicality of our method.

II. BACKGROUND

A. Compressed Sensing MRI

CS-MRI [1] refers to accelerating MRI via *compressed sensing* (CS) [25], which aims to reconstruct the underlying image from a set of subsampled k -space measurements.

1) *Basics*: The common setup of CS-MRI involves the reconstruction of an image $\mathbf{x} \in \mathbb{C}^n$ from its subsampled, noisy k -space measurements

$$\mathbf{y} := \mathbf{M}\mathbf{F}\mathbf{x} + \mathbf{n} \in \mathbb{C}^m \quad (m \ll n), \quad (1)$$

where $\mathbf{F}$ is the Fourier transform, $\mathbf{M} \in \{0, 1\}^{m \times n}$ is the subsampling matrix with $\mathbf{m} \in \{0, 1\}^n$ denoting its subsampling pattern, and $\mathbf{n} \in \mathbb{C}^m$ is the complex measurement noise. For parallel imaging MRI, the measurements are collected from multiple coils. For the i -th coil, the measurements $\mathbf{y}_i$ can be expressed as

$$\mathbf{y}_i := \mathbf{M}\mathbf{F}\mathbf{S}_i\mathbf{x} + \mathbf{n}_i \in \mathbb{C}^m, \quad (2)$$

where $\mathbf{S}_i$ is the pixel-wise sensitivity map and $\mathbf{n}_i$ is the measurement noise of the i -th coil. For both settings, we refer to $b := \|\mathbf{m}\|_1$ as the sampling budget and $R := \frac{n}{b}$ as the acceleration ratio of the acquisition. Classical CS-MRI enables sampling below the Nyquist-Shannon rate by solving an optimization problem with a regularizer that leverages the structure of MRI images [2], [26], [27], [28].

2) *Subsampling Patterns*: Subsampling patterns, or masks, in traditional CS-MRI are often generated randomly or hand-crafted to have a point spread function (PSF) with low coherence, which leads to better reconstruction performance according to the CS theory. Popular subsampling patterns include the 2D variable density [1], bidirectional Cartesian [29], Poisson-disc [30], and continuous-trajectory variable density [31], among others [32], [33]. Despite overall effectiveness, these subsampling patterns are designed for generic image reconstruction and not optimized for any specific body part and diagnostic purpose. Therefore, these patterns may lead to suboptimal performance for downstream tasks where specific anatomical or pathological information is relevant.

3) *DL-Based Reconstruction*: Recently, DL methods have achieved state-of-the-art performance on CS-MRI reconstruction. One line of work combines data-driven priors with model-based iterative reconstruction (MBIR) [3], [4], [34], [35]. Another line of work learns a model-free reconstruction network via end-to-end training [36], [37], [38], [39], [40]. A third line of work, known as deep unrolling (DU), combines the characteristics of MBIR and end-to-end training [6], [8], [41], [42], [43], [44], [45], [46], [47], [48]. The idea is to “unroll” an iterative optimization procedure into a cascade of mappings and train these mappings end-to-end so that they can gradually map a low-resolution input image to a high-quality output reconstruction. Inheriting the advantage of both MBIR and end-to-end learning, these methods exhibit state-of-the-art performance on CS-MRI reconstruction. In this paper, we use a specific kind of unrolled network called E2E-VarNet [43] as part of our framework due to its strong performance on the large-scale fastMRI dataset [49].

B. Reconstruction-Oriented Co-Design

The success of DL methods in CS-MRI reconstruction motivates the idea of jointly optimizing acquisition together with reconstruction via end-to-end training. Recently, there has been a rapidly growing literature on optimizing a parameterized sampling strategy jointly with a CNN reconstructor [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [50]. These methods have different architectural designs and applicable scenarios, but all rely on the differentiable nature of neural networks to optimize the reconstruction accuracy over the choice of k -space measurements. The learned subsampling pattern and reconstruction network are thus specific to the dataset. The end-to-end training enables synergistic cooperation between the learned subsampling pattern and reconstructor, achieving state-of-the-art reconstruction performance. From a task perspective, however, having a reconstruction is not the end of the workflow. These methods rely on either human evaluation, a traditional task prediction algorithm, or a CNN for task predictions, which are out of the scope of these papers.

C. Task-Oriented Co-Design

Recent work has investigated the co-design idea in the context of limited tasks beyond full-FOV reconstruction, such as physical parameter estimation [50], [51], [52] and segmentation [53], [54], [55], [56], [57]. Using task-specific loss functions

in their training procedures, these proposed methods demonstrate stronger task performance than methods trained by a reconstruction-only loss. Most of these proposed approaches leave either subsampling or prediction as a pre-determined fixed module, and focus on co-designing the other modules [50], [51], [52], [53], [54]. On the other hand, the authors of [55], [56] proposed to jointly optimize all three steps, and investigated a brain segmentation task using a U-Net reconstructor and predictor. Although these methods show the potential of extending co-design beyond reconstruction, they are each fine-tuned for one particular task, do not easily accommodate different types of data (e.g., multi-coil), and have not been demonstrated on real out-of-distribution datasets. The most relevant work to ours in the literature is a concurrent work by Wang et al. [57], in which the authors presented a thorough investigation of optimizing the entire CS-MRI pipeline for various segmentation problems. In this work, we cast a wider net for the task-specific CS-MRI co-design problem. In particular we demonstrate our unified framework for designing generalized CS-MRI pipelines, TACKLE, on three different tasks beyond full FOV reconstruction. TACKLE performs robustly on this broad range of tasks and experiments, and is implemented and tested on a Siemens scanner.

III. METHOD

Fig. 2 illustrates the architecture of TACKLE. As a co-design CS-MRI method, TACKLE jointly optimizes the sampler, retriever, and predictor for a task-dependent loss. In the following subsections, we describe each module in order and more implementation details can be found in Supplement II.

A. Sampler

We consider 2D Cartesian subsampling patterns, i.e. $\mathbf{m} \in \{0, 1\}^n$. We follow [9], [15], [58] to model the subsampling strategy as the element-wise Bernoulli distribution with a probability vector $\mathbf{p} \in [0, 1]^n$, i.e. $\mathbf{m}_i \sim \text{Bern}(\mathbf{p}_i)$. To learn the optimal sampling probabilities, we follow the sampler design of [9]. We optimize a set of parameters $\mathbf{q}_i$ that first give us a set of probabilities $\tilde{\mathbf{p}}_i := \text{Sigmoid}(\mathbf{q}_i)$. We then rescale $\tilde{\mathbf{p}}$ to obtain a probabilistic sampling mask $\mathbf{p}$ that would result in b measurements in expectation via Bernoulli sampling:

$$\mathbf{p} = \begin{cases} \frac{\alpha}{\beta} \tilde{\mathbf{p}} & \text{if } \beta \geq \alpha \\ \mathbf{1} - \frac{1-\alpha}{1-\beta} (\mathbf{1} - \tilde{\mathbf{p}}) & \text{otherwise} \end{cases}$$

where $\alpha := \frac{b}{n}$, $\beta := \frac{\|\tilde{\mathbf{p}}\|_1}{n}$, and $\mathbf{1}$ is the all-one vector. During training, the sampler draws a k -space sampling mask $\mathbf{m}$ by sampling $\mathbf{m}_i \sim \text{Bern}(\mathbf{p}_i)$. We repeatedly sample $\mathbf{m}$ until $\|\mathbf{m}\|_1 \approx b$ under a small tolerance. This sampling process encourages exploration of different patterns and ensures the sampling patterns approximately satisfy the budget constraint. Since the sampling process is not differentiable, we use a straight-through estimator to overcome the non-differentiability [59]. During testing, we set the top b indices of $\mathbf{p}$ with the highest probabilities to 1 (sample) and others to 0 (not sample). This binarization guarantees that the sampling mask strictly satisfies the sampling budget constraint

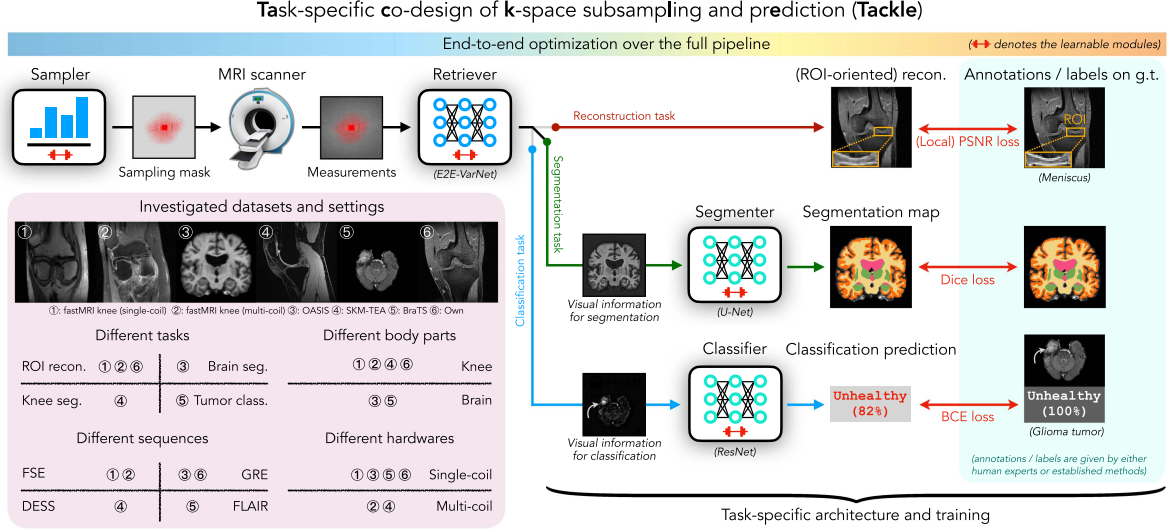


Fig. 2. Block diagram of the proposed framework TACKLE and a summary of the investigated datasets and settings. TACKLE uses a task-specific loss to jointly optimize a sampler, retriever, and an optional predictor, ranging from scanner-level sampling to human-level diagnosis. A summary of the investigated settings is presented in the bottom left panel. FSE, GRE, DESS, and FLAIR stand for fast spin echo, gradient echo, double-echo steady-state, and fluid-attenuated inversion recovery, respectively. We comprehensively investigate multiple CS-MRI tasks on a variety of common MRI settings with six datasets.

and all slices of a volume share the same sampling mask. We also allocate $1/8$ of the sampling budget for the low-frequency region around the DC component, which we refer to as the pre-select region. The pre-selected measurements provide auto-calibration signals (ACS) for multi-coil reconstruction and stabilize the training of some baselines. Therefore, we include the pre-select region for all experiments for consistency. More discussion on this can be found in Supplement II.C. We denote the sampler as S_q where q is the vector of learnable parameters.

B. Retriever

After acquiring measurements, we employ a retriever to extract visual information from noisy and subsampled k -space measurements. We note that we name the module “retriever” instead of “reconstructor” because it is jointly optimized with the downstream predictor for non-reconstruction tasks. Hence, the retriever should not be interpreted as a reconstructor as its output may not be a typical “reconstruction” in terms of point-wise accuracy. We denote the retriever as R_θ where θ is its weights. We select the E2E-VarNet [43] since it is a model-based DU architecture that combines forward model and learning prior, and achieves excellent performance on CS-MRI reconstruction [49]. E2E-VarNet also accommodates multi-coil k -space data with its ability to estimate coil sensitivity maps. Specifically, our E2E-VarNet retriever operates in k -space and consists of 12 refinement steps, each of which includes a U-Net [60] with independent weights from each other. For each U-Net, we use the standard architecture with the following parameters: 2 input and output channels, 18 channels after the first convolution filter, 4 average down-pooling layers, and 4 up-pooling layers. The final output layer of the retriever is an inverse Fourier transform followed by a root-sum-squares reduction for each pixel over all coils. The output of the retriever is a batch of magnitude images.

For reconstruction tasks, a loss function will be directly applied to the output. For non-reconstruction tasks, the output will be fed into an additional predictor module described in the next section.

C. Task-Specific Design: Predictor and Loss Function

We demonstrate TACKLE on three tasks that together represent a gradual progression from generic full-FOV reconstruction to clinically relevant tasks.

1) *ROI-Oriented Reconstruction*: For many MRI scans, only a small region of the FOV is relevant to the reader, so we define a task where we aim to maximize reconstruction quality around that region. In contrast to the full-FOV reconstruction task, the reconstruction accuracy in this task is only measured over the region-of-interest (ROI) of each image instead of the entire FOV. We hereafter refer to this task as *ROI-oriented reconstruction*. This task is a first step from generic full-FOV reconstruction to more specific downstream tasks in CS-MRI.

There is no predictor for this reconstruction task, and the output of the retriever will directly be used for evaluation. The evaluation metric we use is the local peak signal-to-noise ratio (PSNR), which is the PSNR within the ROI of an underlying image x . Let $\mathcal{R}_x$ be the set of indices i that are within the ROI of x . Note that $\mathcal{R}_x$ varies from one image x to another. We define the local PSNR within the ROI as

$$\text{LocalPSNR}(\hat{x}, x; \mathcal{R}_x) := 10 \log_{10} \frac{\max(x)^2}{\text{LocalMSE}(\hat{x}, x; \mathcal{R}_x)} \quad (3)$$

where $\text{LocalMSE}(\hat{x}, x; \mathcal{R}_x) := \frac{1}{|\mathcal{R}_x|} \sum_{i \in \mathcal{R}_x} (\hat{x}_i - x_i)^2$ and $\max(x)$ is the largest pixel value of x . We optimize our model for the local reconstruction quality using $\mathcal{L}_{\text{ROI}}(\hat{x}, x) := -\text{LocalPSNR}(\hat{x}, x; \mathcal{R}_x)$ as the training loss.

2) *Tissue Segmentation*: For this task, we aim to predict segmentation maps of different body tissues. Accurately segmenting a tissue from the rest of the organ provides important anatomical and pathological information [61], [62], [63]. Conventional segmentation workflow involves human evaluation and traditional algorithms, which often require standard reconstructions of certain contrasts as input [64]. On the contrary, TACKLE does not require reconstruction as a necessary intermediate step, and is optimized for segmentation performance in an end-to-end fashion.

We include an additional predictor P_ϕ with weights ϕ subsequent to the retriever. We choose the U-Net architecture due to its ability of solving medical image analysis tasks [60], [65], [66]. The specific parameters are: 1 input channel, c output channels (where c is the number of segmentation classes), 64 channels after the first convolution filter, 4 average down-pooling layers, and 4 up-pooling layers.

We use the Dice score [67], [68], [69] as the evaluation metric. The Dice score measures the degree of overlap between two segmentation maps and takes a value between 0 (no overlap) and 1 (perfect overlap). During training, we employ the Dice loss $\mathcal{L}_{\text{seg}}(\hat{z}, z) := 1 - \text{DiceScore}(\hat{z}, z)$. For both training and evaluation, we apply a Softmax function across all the classes for each pixel and then calculate the Dice loss/score. During the evaluation, we apply an additional binarization step where we set the class with highest value after Softmax as 1 and others as 0. In this way, we assign each pixel of the predicted segmentation map $\hat{z}$ to exactly one class.

3) *Pathology Classification*: The third task we consider is to determine whether a potential pathology exists in an MRI image, such as a suspected tumor. Using algorithms to automatically analyze MRI scans could lead to improved diagnosis accuracy in clinical practice [24]. We formulate this task as a binary image classification problem, where the negative class means the underlying image x does not contain any pathology lesion and the positive class means it does contain a lesion. Through this proof-of-concept classification task, we go beyond pixel-level problems and show the benefit of task-specific co-design for solving an image-level problem.

Similar to the segmentation task, we include an additional predictor in the pipeline, which we also denote as P_ϕ to simplify notations. Specifically, we choose the ResNet [70], which is an established architecture for computer vision tasks, especially image classification. We use the standard ResNet18 architecture except for using 1 input channel and 2 output dimensions.

We use the binary cross entropy (BCE) as the loss function for this classification task, $\mathcal{L}_{\text{class}}(\hat{z}, z) := \text{BCE}(\hat{z}, z)$. For evaluation metrics, we consider both the classification accuracy ($\text{ClsAcc} := \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$) and the F_1 score ($F_1 \text{ score} := \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$) where TP, TN, FP, and FN are the number of True Positive, True Negative, False Positive, and False Negative, respectively. The classification accuracy is more interpretable, while the F_1 score is more robust to class imbalance. So we include both metrics for a more comprehensive evaluation.

D. Training Procedure

We summarize the training objective for each task as follows:

- ROI-oriented reconstruction:

$$\min_{q, \theta} \mathcal{L}_{\text{ROI}}(R_\theta(S_q \odot k), x)$$

- Segmentation or classification:

$$\min_{q, \theta, \phi} \mathcal{L}_{\text{seg./class.}}(P_\phi(R_\theta(S_q \odot k)), z)$$

where $k \in \mathbb{C}^n$ contains all k -space measurements of x .

When performing end-to-end training over multiple stages, we empirically observed that a model trained from scratch tends to run into either optimization (hard to train) or generalization (unable to generalize) issues. Some prior works address these problems using a hybrid of reconstruction and task-dependent loss [51], [52], [54], [55], [56]. This approach requires tuning a weight parameter that balances the two losses. We adopt an alternative approach that avoids tuning this additional parameter. Specifically, we first train the sampler and retriever jointly with a full-FOV PSNR loss until convergence:

$$\min_{q, \theta} \mathcal{L}_{\text{FOV}}(R_\theta(S_q \odot k), x)$$

where $\mathcal{L}_{\text{FOV}}(\hat{x}, x) := -\text{PSNR}(\hat{x}, x)$. We refer to this as the pre-training step in later sections. With the weights learned for the sampler and retriever, we then add the predictor (initialized with random weights) into the framework and fine-tune all three components. We find that the pre-training step allows the model to better learn task-specific strategies, as demonstrated by an ablation study in Section VI-B. This training procedure mimics the training of foundation models in state-of-the-art language and vision models, which are first pre-trained on a general task and then fine-tuned for more specific tasks. Similar procedures can be found in other task-specific co-design papers, such as [53], [54].

IV. EXPERIMENTS ON LARGE-SCALE DATASETS

We first demonstrate the effectiveness of our framework on the three considered tasks using large-scale datasets. We categorize all the investigated datasets and settings in the bottom left panel of Fig. 2. For each task, we demonstrate that the proposed task-specific co-design framework achieves better performance than baselines that separately design reconstruction and prediction. We abbreviate different variants of the proposed method and baselines in the following way based on their task and training procedure:

Task	Training procedure and loss	Notation example
(ROI) recon.	S&R: PSNR loss S&R: local PSNR loss (w/ warm-up)	LP+UN _{FOV} TACKLE _{ROI}
Tissue seg.	S&R: PSNR loss → P: Dice loss S&R&P: Dice loss (w/ warm-up)	PD+UN _{recon.} TACKLE _{seg.}
Patho. class.	S&R: PSNR loss → P: BCE loss S&R&P: BCE loss (w/ warm-up)	LOUPE _{recon.} TACKLE _{class.}

S: sampler, R: retriever, P: predictor

→: separate training stages for reconstruction and prediction

To clarify, the subscript “recon.” for the segmentation and classification methods means that the sampler and retriever are trained for full-FOV reconstruction, and a predictor is subsequently trained for the downstream task with the sampler and retriever fixed. This is equivalent to training a predictor with

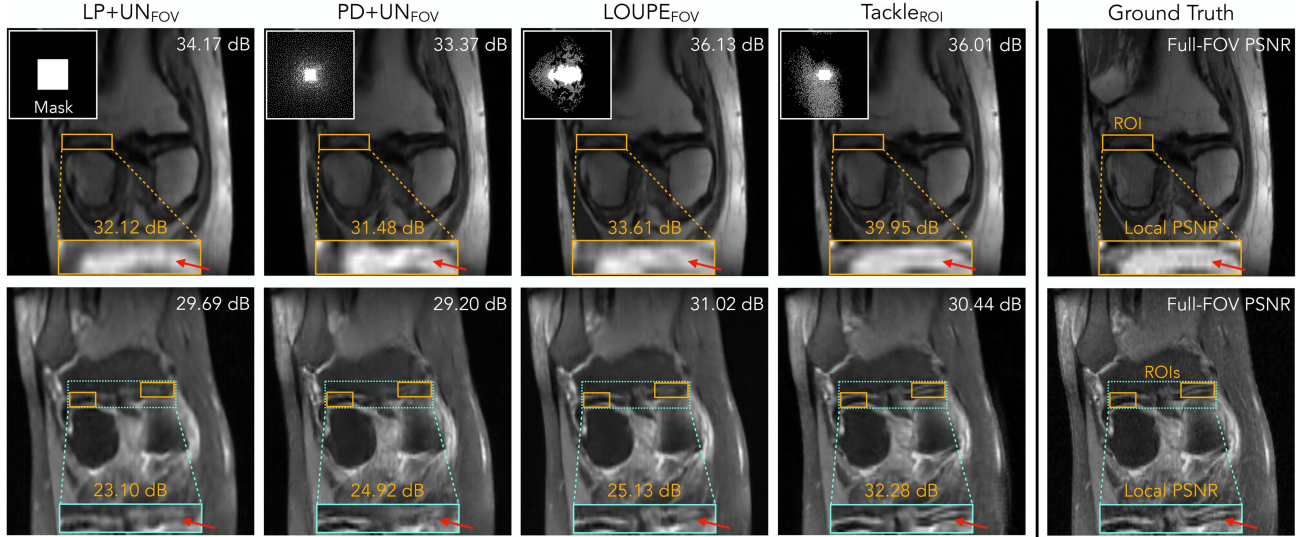


Fig. 3. Visual examples of two Meniscus Tear samples reconstructed by different methods in the $16\times$ acceleration single-coil setting. For each reconstruction, the full-FOV PSNR is labeled in white, and the local PSNR for the ROI is in orange. Note how TACKLE_{ROI} recovers the structure and details of the ROI more accurately than the two baselines, as indicated by the red arrows. The better recovery of TACKLE_{ROI} over the ROI leads to a more accurate diagnosis of the Meniscus Tear. We emphasize that the location of the ROI is not an input to any of these models and is only used for evaluating the accuracy of each method on the region that contains the pathology.

TABLE I
COMPARISON OF AVERAGE TEST LOCAL PEAK SIGNAL-TO-NOISE RATIO (LOCAL PSNR) IN DECIBEL (dB) WITHIN MENISCUS TEAR ROIS

Data	R	LP+UN _{FOV}	PD+UN _{FOV}	LOUPE _{FOV}	TACKLE _{ROI}
Single-coil	8	26.95	28.23	30.32	34.04
	16	25.16	26.05	27.32	31.54
Multi-coil	8	27.55	32.68	34.88	40.65
	16	26.02	30.00	31.79	37.89

the reconstructed images by these methods as input for the downstream task.

A. ROI-Oriented Reconstruction

Dataset and setup: For the ROI-oriented reconstruction task, we use the images and raw single- and multi-coil k -space data from the fastMRI+ knee dataset [49], [71], which contains bounding box annotations for knee pathologies. Specifically, we investigate the most common knee pathology in the dataset called “Meniscus Tear” (MT). Each image x in the dataset contains at least one rectangular bounding box annotation $\mathcal{R}_x$, which is drawn to include all the pathology but exclude the normal surrounding anatomy [71]. Therefore the local image quality within each bounding box (i.e. ROI) is more indicative of the quality for pathology assessment than a metric over the entire FOV. We emphasize that the location of the bounding box $\mathcal{R}_x$ varies sample by sample and is never an input to any method during inference. $\mathcal{R}_x$ is only used for calculating the training loss and evaluating the local PSNR during test time according to (3). Hence, the local PSNR performance reflects the quality of reconstructions by different methods for assessing the considered pathological lesions in the ROIs.

Baselines: We compare TACKLE_{ROI} with three full-FOV reconstruction-oriented baselines.

- **LOUPE_{FOV}:** Proposed in [9], LOUPE_{FOV} jointly optimizes a sampler and a residual U-Net reconstructor.
- **Low-pass + U-Net_{FOV} (LP+UN_{FOV}):** substitute the sampler in LOUPE_{FOV} with a fixed low-pass filter sampling pattern.
- **Poisson-disc + U-Net_{FOV} (PD+UN_{FOV}):** substitute the sampler in LOUPE_{FOV} with a Poisson-disc sampling pattern drawn from a variable density distribution and generated with the sigpy.mri.poisson function in the SigPy package.¹

Results: We compare the average local PSNR of our method and other baselines over the test set in Table I. For all settings, TACKLE outperforms other baselines designed for full-FOV reconstruction by at least 3 dB, indicating a significant improvement of image quality within the ROI.

In Fig. 3, we provide example reconstructions by our method and three baseline methods. For each reconstruction, a zoom-in on its ROI is provided on the bottom with the corresponding local PSNR value labeled above in orange, and its full-FOV PSNR is labeled on the top right corner of each reconstruction. As shown in the ground truth of the MT example, a meniscus tear is indicated by a streak (dark in the top row and bright in the bottom row) that is present on the meniscus (bright in the top row and dark in the bottom row), as indicated by the red pointers. To accurately detect the existence and assess the severity of a meniscus tear, a reconstruction should clearly show the boundaries of the meniscus and details of the tear. However, the ROIs of both LP+UN_{FOV} and LOUPE_{FOV} reconstructions contain significant reconstruction artifacts that disguise the tear

¹<https://github.com/mikgroup/sigpy>

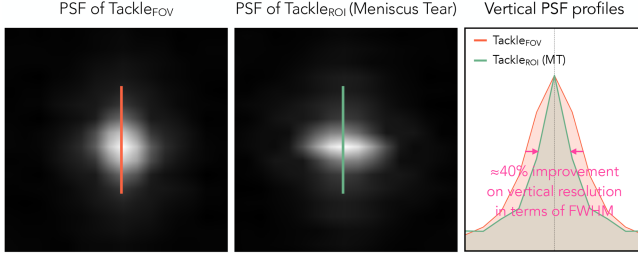


Fig. 4. Comparison of a subsampling PSF optimized for full-FOV reconstruction and another optimized for the reconstruction of meniscus tear (MT) ROIs. Optimizing for MT ROI reconstruction leads to around 40% improvement on the vertical resolution in terms of the *full width at half maximum (FWHM)*, as shown by the PSF profiles in the bottom panel. This improved vertical resolution leads to a better reconstruction of the meniscus that has horizontal anatomy.

(see the red arrows). On the other hand, TACKLE_{ROI} preserves the details of the tear and contains fewer artifacts than the baselines, providing a more accurate ROI reconstruction with a higher diagnostic value.

In Supplement IV we also include a validation of TACKLE_{ROI} on images that either are healthy or contain pathologies other than the meniscus tear. Although TACKLE_{ROI} is not designed to generalize across different pathologies, we empirically find that TACKLE_{ROI} still yields high-fidelity reconstructions for out-of-distribution images so that the pathologies on these images remain detectable. We also find that TACKLE_{ROI} generalizes consistently across the three acceleration ratios (4×, 8×, and 16×) for the fastMRI+ dataset.

Discussion: Enhancing local ROIs for MRI may seem counter-intuitive, because the acquisition happens in k -space; each frequency measurement in theory corresponds to the entire FOV. Here we understand the feasibility via a PSF analysis. Consider the zero-filled reconstruction $\tilde{x}$ from some (noiseless) single-coil k -space data:

$$\tilde{x} := F^{-1}(\mathbf{m} \odot (F\mathbf{x})) = (F^{-1}\mathbf{m}) * \mathbf{x}$$

where $*$ denotes convolution and the second equality holds due to the Fourier convolution theorem. Here, $F^{-1}\mathbf{m}$ is the PSF of the subsampling mask $\mathbf{m}$ and determines the resolution of the CS-MRI system. We visualize the PSF of a sampling mask trained for full-FOV reconstruction and another trained for meniscus tear (MT) ROIs reconstruction with the same sampling budget in Fig. 4. We plot the PSF profiles in the vertical direction around the main lobes. The PSF learned for MT ROI reconstruction has around 40% improvement in vertical resolution in terms of *full width at half maximum (FWHM)* of the PSF profiles. Since MT ROIs contains the thin horizontal anatomy of the meniscus, it makes sense that the learned subsampling pattern has a narrower PSF profile (and thus higher resolution) in the vertical direction. This comparison demonstrates that the improvement on ROIs is partly due to the capability of our model to optimize the subsampling PSF for local ROI anatomy via co-design. This is particularly beneficial when there is a mismatch between the optimal subsampling PSF for full-FOV reconstruction and that for ROI reconstruction due to directional anatomical structure, which is the case for MT ROI reconstruction.

TABLE II
COMPARISON OF AVERAGE TEST DICE SCORE ON THE SKM-TEA DATASET [72] FOR SEGMENTING FOUR KNEE TISSUES UNDER DIFFERENT ACCELERATION RATIOS (R)

R	PD+UN _{recon.}	LOUPE _{recon.}	SemuNet	TACKLE _{recon.}	TACKLE _{seg.}
16	0.7843	0.7888	0.8108	0.8232	0.8532
64	0.7486	0.6715	0.7741	0.8145	0.8357

B. Knee Tissue Segmentation²

Dataset and setup: This study involves segmenting four types of knee tissues: the patellar cartilage, the femoral cartilage, the tibial cartilage, and the meniscus. We use the *Stanford Knee MRI with Multi-Task Evaluation (SKM-TEA)* dataset [72], which contains pixel-level segmentation maps of the four tissues. Specifically, we use the raw 3D multi-coil k -space measurements of knee images and take 1D inverse Fourier transform along the left-to-right direction to obtain 2D k -space of sagittal slices. We train each method to minimize the Dice loss until convergence and select the model with the highest Dice score on the validation set.

Baselines: We compare TACKLE_{seg.} with four baselines.

- *LOUPE_{recon.}:* LOUPE_{recon.} is a baseline based on LOUPE_{FOV}. We first train a LOUPE_{FOV} model for the full-FOV reconstruction task and then use the reconstructed images to separately train a segmentation network.
- *Poisson-disc + U-Net_{recon.} (PD+UN_{recon.}):* same as LOUPE_{recon.} except that the sampler is fixed to be a Poisson-disc sampling mask.
- *TACKLE_{recon.}:* same as LOUPE_{recon.} except for using the proposed architecture of TACKLE.
- *SemuNet:* Proposed in [56], SemuNet uses a hybrid of ℓ_1 reconstruction loss and cross-entropy segmentation loss.

Results: We provide a quantitative comparison in Table II and a boxplot comparison in Fig. 5. Within the rectangle between each pair of methods in Fig. 5, the top number is the percentage of samples that get improved and the bottom number is the p -value given by the paired samples t-test. With an improved architecture, TACKLE_{recon.} already outperforms the other baselines. Nevertheless, the segmentation-oriented method TACKLE_{seg.} achieves even better performance on both 16× and 64× accelerations. TACKLE_{seg.} also significantly outperforms SemuNet on both acceleration ratios and has a much smaller performance drop from 16× to 64× than SemuNet, indicating that the proposed approach is more robust to high acceleration ratios. We further provide some visual examples in Fig. 6. The first row visualizes the input of the predictor by different methods, where each image is labelled by its PSNR value on the top right corner. The last row shows the predicted segmentation maps by different methods, where each prediction is labelled by its Dice score on the top right corner. The blue arrows point out the locations where TACKLE_{seg.} provides more accurate reconstructions than other reconstruction-oriented baselines. We also provide a zoom-in on the region that contains the segmented tissues in the second row.

²See brain segmentation results in Supplement I.

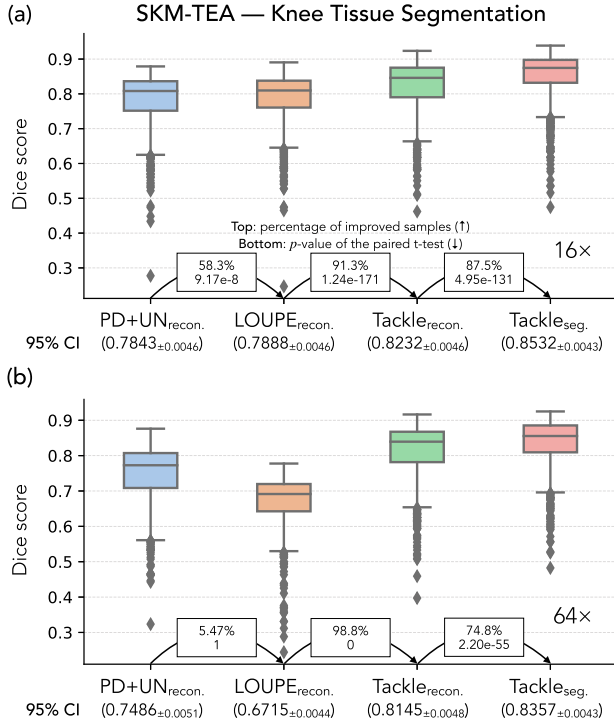


Fig. 5. Box plots of the knee tissue segmentation results under $16\times$ (a) and $64\times$ (b). Within the rectangle between each pair of methods, the top number is the percentage of samples that get improved and the bottom number is the p -value given by the paired samples t -test. A higher percentage and lower p -value indicate a more significant improvement. We also provide the 95% confidence intervals for all methods below their names. For both acceleration ratios, TACKLE_{seg} outperforms other baselines in terms of all the statistical measures.

Discussion: We note that TACKLE_{seg} learns an intermediate feature map as the input to the predictor, which circumvents a typically “good” reconstruction; it is interesting how the retriever produces an image where different knee tissues to be segmented have distinctive textures, which are easy to distinguish both from the background and from each other. Even though this feature map is not a typical “reconstruction” in terms of point-wise accuracy, it still accurately localizes the anatomy of the tissues to be segmented. We highlight that TACKLE_{recon} provides a high-fidelity reconstruction of the entire FOV with a PSNR of 33.00 dB, which demonstrates that our model is well capable of doing the reconstruction task accurately. However, TACKLE_{seg} still outperforms TACKLE_{recon} in terms of segmentation performance in Fig. 6 and on average over the dataset in Table V (see Section VI-A for more details). This observation demonstrates that finding the most accurate full-FOV reconstruction does not necessarily lead to the optimal result on the considered segmentation task.

C. Pathology (tumor) Classification

Dataset and setup: In this section, we demonstrate the effectiveness of the proposed method at detecting the existence of gliomas, a common type of brain tumors in adults. We use the images acquired by the FLAIR sequence in the Multimodal Brain

TABLE III
COMPARISON OF AVERAGE TEST ACCURACY ON THE PATHOLOGY CLASSIFICATION TASK UNDER DIFFERENT ACCELERATION RATIOS (R)

Metric	R	PD+UN _{recon.}	LOUPE _{recon.}	TACKLE _{recon.}	TACKLE _{class.}
Cls. acc.	16	0.9016	0.9024	0.9062	0.9159
	64	0.8809	0.8930	0.9054	0.9136
F_1 score	16	0.8853	0.8846	0.8929	0.9039
	64	0.8628	0.8768	0.8910	0.8992

Tumor Image Segmentation Benchmark (BRATS) dataset [24]. To obtain an image-level label of the existence of a tumor, we aggregate the pixel-level peritumoral edema (ED) segmentation annotations in the BRATS dataset by checking whether there exists any positive pixel in the segmentation map: negative (healthy) means there is no ED pixel, while positive (unhealthy) means there is at least one ED pixel. We simulate the single-coil k -space data for each image by taking the Fourier transform of the image and adding complex additive white Gaussian noise (AWGN), according to the forward model in (1). The standard deviation of the noise for each image is 0.05% of the magnitude of the DC component. We train all models using the BCE loss and evaluate them using the classification accuracy and F_1 score as described in Section III-C3.

Baselines: We compare the proposed method TACKLE_{class} with the first three baselines as in Section IV-B except that the predictor of each baseline is subsequently trained for pathology classification rather than tissue segmentation (with input images optimized for full-FOV reconstruction). We do not include SemuNet here because it is originally proposed for the segmentation task only.

Results: In Table III, we compare the classification-oriented method, TACKLE_{class}, with reconstruction-oriented baselines, and find that TACKLE_{class} achieves higher classification accuracy under both performance metrics. Specifically, TACKLE_{class} outperforms the existing reconstruction-oriented baseline LOUPE_{recon} by around 2% in the extreme $64\times$ accelerated acquisition scenario. Both variants of TACKLE maintain competitive performance under the highly accelerated setting ($R=64$), while PD+UN_{recon} and LOUPE_{recon} suffer from significant performance degradation. Note that TACKLE_{class} outperforms TACKLE_{recon} by more than 0.8% in both cases, despite having the same architecture. We also visualize and compare the classification performance of TACKLE_{class} and LOUPE_{recon} under $16\times$ acceleration in Fig. 7, using confusion matrices. The results show that TACKLE_{class} has substantially fewer false negatives (bottom left) and a higher overall accuracy compared to LOUPE_{recon}.

V. VALIDATION ON AN EXPERIMENTALLY COLLECTED OUT-OF-DISTRIBUTION DATASET

In practice, creating a large well-annotated training set for a specific task can be very time-consuming or even infeasible. To demonstrate the immediate benefit of our method in a real-world setting, we conduct a validation of TACKLE on the ROI-oriented reconstruction task using experimentally collected data that is out of the distribution of the training data. Specifically, we train a

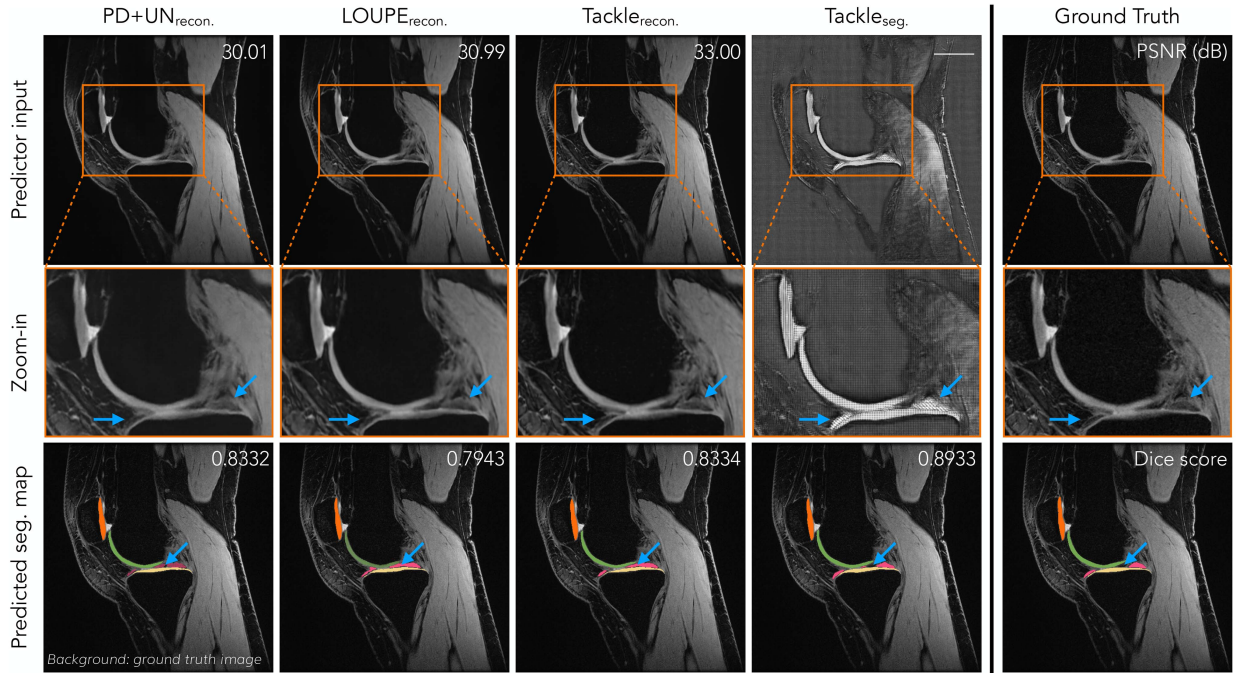


Fig. 6. Comparison of segmentation results under $16\times$ acceleration on one sample of the SKM-TEA dataset. We show the input of the predictor in the first row, a zoom-in on the region that contains the tissues to be segmented in the second row, and the output of the predictor in the third row. Note that TACKLE_{seg.} circumvents the typical “reconstruction” in terms of point-wise similarity with the ground truth image. Instead, it learns a feature map that accurately localizes the anatomy, leading to better segmentation prediction than other baselines both for this sample and on average over the test set (Table II).

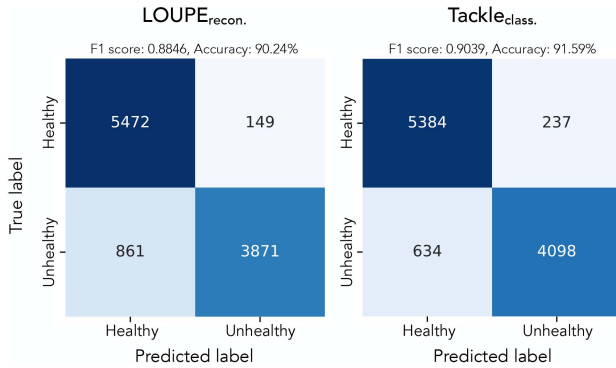


Fig. 7. Confusion matrices of the classification results by LOUPE_{recon.} and TACKLE_{class.}. Overall, TACKLE_{class.} achieves higher accuracy in terms of both classification accuracy and F_1 score than LOUPE_{recon.}. TACKLE_{class.} also has a significantly lower number of false negatives (bottom left) compared to LOUPE_{recon.}, which could lead to more patients receiving early treatment.

TACKLE model on a large-scale dataset (fastMRI in this case) and directly test it on raw k -space data collected by *different hardware using a different type of sequence* from that of the training. Even without extra fine-tuning or test-time optimization, the learned ROI-specific model provides improved reconstructions on meniscus ROIs. In the following subsections, we present the details of this experiment.

Data acquisition and processing: Two subjects were scanned at the Massachusetts General Hospital in accordance with institutional review board guidelines. Their right knees were scanned by a 3D-encoded Cartesian gradient-echo sequence with a 3

Tesla MRI scanner (Model: Skyra; Siemens Healthcare, Erlangen, Germany) and a single-channel extremity coil. To implement the 2D subsampling pattern in the coronal plane, we used a transversal orientation with the frequency encoding direction (k_x) pointing into the knee cap (anterior-posterior), so that the two phase encoding directions were left-right (k_y) and superior-inferior (k_z), respectively. The acquisition parameters were as follows: TE/TR=4.8/9.1 ms, FOV=192 × 192 × 192 mm³, resolution = 1 × 1 × 1 mm³, flip angle = 10°. The total acquisition time of obtaining the fully sampled data for each subject was 5 minutes and 35 seconds. The raw k -space data had the shape of 192 × 192 × 192 ($k_x \times k_y \times k_z$). We applied the 1D inverse Fourier transform along k_x for downstream processing. Specifically, we took the middle 40 slices of each volume and annotated bounding boxes around the meniscus region using an image labelling tool.³ Efforts were made such that the locations and sizes of the bounding boxes roughly match those in the fastMRI MT dataset. We emphasize that these bounding boxes are *only* for the purpose of measuring the accuracy of different models on reconstructing the meniscus region. The locations of the annotated ROIs are *not* the input to any of the tested models.

Generalization gaps: There are multiple generalization gaps between the training (fastMRI single-coil data) and test data:

- **Different hardware:** The acquired data are collected directly with a single-channel extremity coil, while the training data are simulated from k -space data collected by multi-channel receiver coils [73].

³<https://github.com/heartexlabs/labelImg>

TABLE IV
AVERAGE RECONSTRUCTION ACCURACY ON THE EXPERIMENTALLY
COLLECTED DATASET UNDER $4\times$ ACCELERATION (TOP: FULL-FOV RECON.;
BOTTOM: ROI-ORIENTED RECON.)

Full-FOV recon.	PD+TV _{FOV}	LOUPE _{FOV}	TACKLE _{FOV}	TACKLE _{ROI}
PSNR (dB)	27.94	28.00	28.70	28.18
ROI recon.	PD+TV _{FOV}	LOUPE _{FOV}	TACKLE _{FOV}	TACKLE _{ROI}
Local PSNR (dB)	24.45	24.67	25.16	25.72

indicates the variant of TACKLE with matching training and evaluation metrics

- *Different sequence and resolution*: The acquired data are given by a gradient-echo sequence with 1 mm isotropic resolution, while the training data are given by a spin-echo sequence with 0.5 mm in-plane resolution [73].
- *Different distribution of the ROI anatomy*: The acquired data are collected from two subjects whose menisci are healthy and have no tear, while the ROIs in the training data contain meniscus tears.

Despite these generalization gaps, TACKLE_{ROI} works robustly and leads to both numerical and visual improvement.

Baselines: In this section, we compare TACKLE_{ROI} with the following baselines under $4\times$ acceleration.

- *Poisson-disc + Total Variation_{FOV} (PD+TV_{FOV})*: The sub-sampling pattern is a Poisson-disc sampling mask generated with the `sigpy.mri.poisson` function in the `SigPy` package.⁴ The reconstruction is obtained by solving a total variation (TV) regularized optimization problem with the Sparse MRI toolbox.⁵
- *LOUPE_{FOV}*: the same LOUPE_{FOV} baseline as in Section IV-A.
- *TACKLE_{FOV}*: a TACKLE model trained for full-FOV reconstruction.
- *LOUPE_{ROI}*: the same architecture as LOUPE_{FOV} but trained for ROI reconstruction following the same training procedure as TACKLE_{ROI}.

Results: We present a quantitative comparison in Table IV for both the full-FOV and ROI-oriented reconstruction tasks. For both tasks, TACKLE outperforms the baselines under the corresponding metric. For each task, we highlight the variant of TACKLE trained for the evaluation metric in green. Our results show that the highlighted variant outperforms the other variant of TACKLE, indicating a tradeoff between full-FOV and ROI reconstruction accuracy.

We further conduct a slice-wise PSNR analysis in Fig. 8. For both histograms, the horizontal axis is the improvement by the respective metric and the vertical axis is the count. We also quantify the significance of the improvements using the paired samples t-test. For the full-FOV reconstruction, TACKLE_{FOV} outperforms LOUPE_{FOV} on *all* 80 slices, giving a highly significant p -value of $3.10e-57$. We then compare TACKLE_{ROI} with the better full-FOV reconstruction method, TACKLE_{FOV}, on the

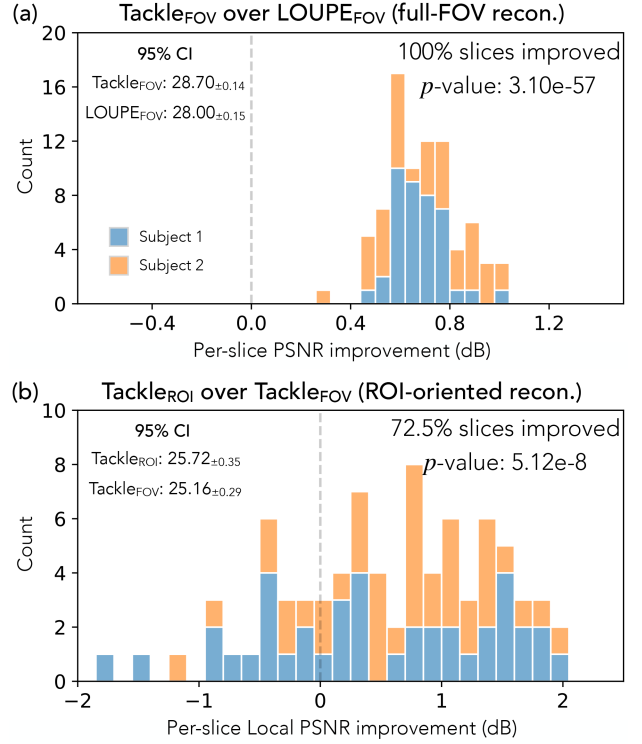


Fig. 8. Slice-wise difference histograms. (a): TACKLE_{FOV} over LOUPE_{FOV} on the full-FOV reconstruction task and (b): TACKLE_{ROI} over LOUPE_{FOV} on the ROI-oriented reconstruction task. The 95% confidence intervals are given in the top left corner of each plot. In both cases, the vast majority of slices are improved and the p -values given by the paired samples t-test are highly significant.

ROI-oriented reconstruction task. Despite having the same architecture, TACKLE_{ROI} still outperforms TACKLE_{FOV} on 72.5% of slices, leading to a p -value of $5.12e-8$, which is also statistically significant. This result indicates that the ROI-oriented model TACKLE_{ROI} indeed provides more accurate ROI reconstructions on this out-of-distribution dataset. We further provide some visual examples in Fig. 9. Below each reconstruction is a zoom-in on the region around the ROI and the error map of the region with respect to the ground truth. TACKLE not only achieves higher PSNR values in both cases but also visually recovers the ROIs with fewer artifacts.

Implementation: Besides the above results based on retrospective subsampling for quantitative comparison, we have also tested the learned sequence on a Siemens 3T MRI Skyra scanner. Specifically, we implement a re-ordering loop that iterates through all the trajectories based on our learned subsampling mask m . The implemented sequence prospectively subsamples in k -space and shortens the scan time from 335 seconds to 84 seconds. In Fig. 10, we compare the reconstruction given by the prospectively subsampling sequence we implement with the reconstruction given by the retrospectively subsampled measurements from the fully sampling sequence. We note that the images labelled as "TACKLE_{ROI} (retrospective)" and "TACKLE_{ROI} (prospective)" are taken by two consecutive

⁴<https://github.com/mikgroup/sigpy>

⁵<https://people.eecs.berkeley.edu/~mlustig/Software.html>

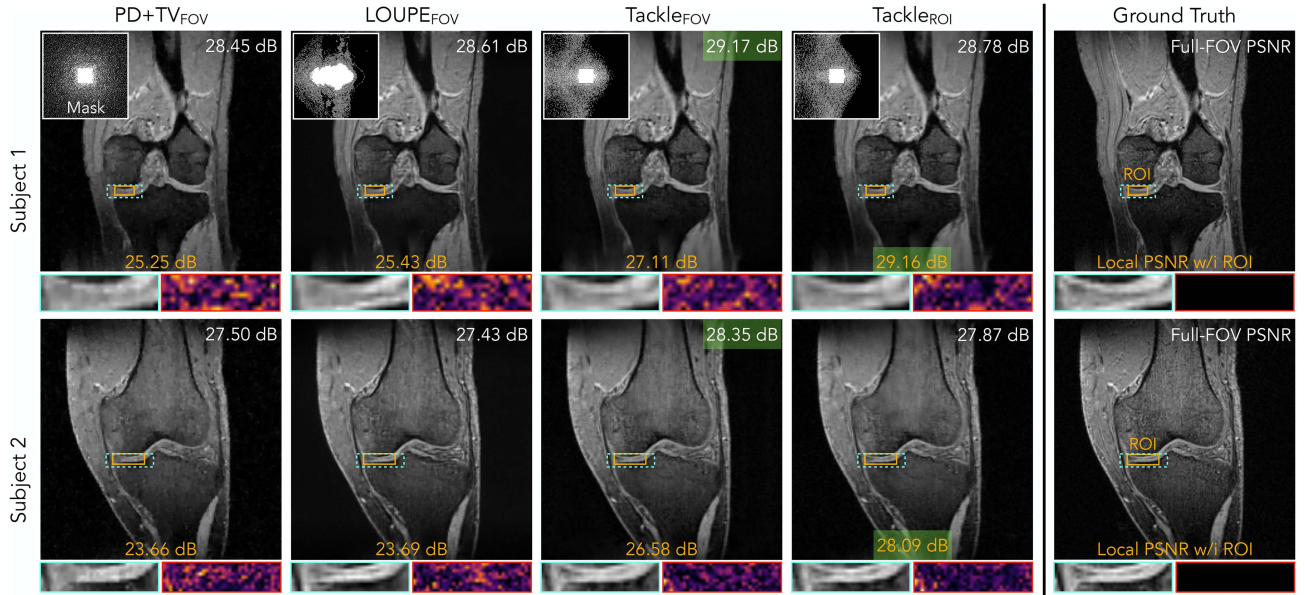


Fig. 9. Reconstruction comparison of two samples in the experimentally collected dataset (top: from subject 1; bottom: from subject 2) by different methods under $4\times$ acceleration. The sampling mask, a zoom-in on the ROI, and the error map are presented for each method. By sampling more frequencies along the vertical direction in k -space, TACKLER_{ROI} has a higher vertical resolution in the image space and thus outperforms other baselines optimized for full-FOV reconstruction on the ROIs with directional anatomical structure.

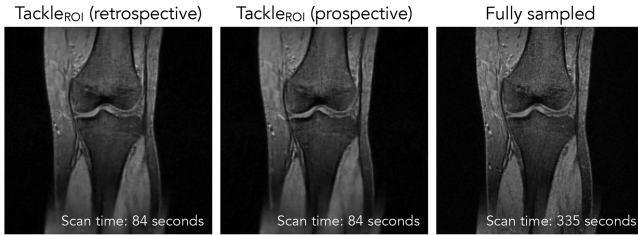


Fig. 10. Reconstruction comparison between the implemented prospective subsampling sequence and the retrospective subsampling sequence. Our learned sequence can be implemented on an MRI scanner and generates images of quality indistinguishable from those recovered from retrospectively sampled data. Compared to the ground truth image, our prospectively subsampled reconstruction recovers important features around the meniscus region, which is the ROI it was trained to enhance.

but separate scans, so there might be some subtle motion between them. Nevertheless, the two images have no significant visual difference, indicating that the improvement we show on retrospective simulations translates into actual improvement in practice. The prospective reconstruction successfully recovers important anatomical features around the meniscus region while only takes a quarter of the scan time compared to the full-sampled image.

VI. ABLATION STUDIES

A. Effectiveness of Co-Design

We evaluate the effectiveness of two aspects of co-design used in the proposed framework: learnable subsampling and task-specific training. In Table V, we compare four variants of

the proposed method that have neither, one, and both aspects of co-design, respectively. The meanings of having or not having each aspect are summarized as follows:

- Learnable subsampling (column 2)
 - $\times$ (*Poisson-disc*): use a Poisson-disc subsampling pattern that is randomly generated and then fixed
 - $\checkmark$: learn the subsampling pattern from data
- Task-specific training (column 3)
 - $\times$: separately optimize retriever and predictor
 - $\checkmark$: jointly optimize retriever and predictor

To eliminate the effect of different network architectures, all four variants have the same retriever (E2E-VarNet) and predictor (U-Net for segmentation and ResNet for classification) architectures. Overall, we find both aspects of co-design beneficial. For the task of reconstructing meniscus tear ROIs, learning the subsampling pattern is particularly helpful. Task-specific training, on the other hand, is more important for the knee segmentation task. Highlighted in cyan, the last row is the full-fledged version of TACKLE, which achieves the best performance for all considered scenarios with both aspects of co-design.

B. Effectiveness of the Proposed Architecture and Training Procedure

The proposed architecture of $\mathcal{T}_\theta$ from measurements $\mathbf{y}$ to prediction $\hat{\mathbf{z}}$ consists of an E2E-VarNet retriever and a U-Net predictor. A natural question is how this architecture compares with a single model-free neural network with a comparable number of parameters that directly maps subsampled measurements to the final prediction. We consider the following comparisons in Table VI:

TABLE V
ABLATION STUDIES ON TWO ASPECTS OF CO-DESIGN FOR ALL THE CONSIDERED TASKS UNDER $16\times$ ACCELERATION

Method	Ablated component		ROI-oriented reconstruction (Local PSNR in dB)		Tissue segmentation (Dice score)		Pathology classification (Cls. acc.) (F_1 score)	
	Learned subsampling	Task-specific training	Single-coil	Multi-coil	Brain [¶]	Knee	Gliomas tumor	
PD+VN _§	✗ (Poisson-disc)	✗	29.91	36.48	0.9257	0.8018	0.9024	0.8871
PD+VN _‡	✗ (Poisson-disc)	✓	30.15	36.51	0.9256	0.8474	0.9072	0.8966
TACKLE _§	✓	✗	30.14	37.53	0.9350	0.8232	0.9062	0.8929
TACKLE _‡	✓	✓	31.54	37.89	0.9395	0.8532	0.9159	0.9039

[‡] indicates full-FOV reconstruction oriented versions of PD+VN and TACKLE
[¶] indicates task-specific versions of PD+VN and TACKLE

¶ see Supplement I

TABLE VI
ABLATION STUDIES ON MODEL ARCHITECTURE AND PRE-TRAINING FOR
NON-RECONSTRUCTION TASKS UNDER $16\times$ ACCELERATION

Ablated component		Tissue segmentation (Dice score)		Pathology classification (F_1 score) (Cls. acc.)	
Arch. of $\mathcal{T}_\theta$	Pre-train	Brain [¶]	Knee	Gliomas tumor	
Predictor only [‡]	✗	0.9005	0.7539	0.8966	0.8788
VN+predictor [‡]	✗	0.9371	0.8163	0.9102	0.8969
VN+predictor [§]	✓	0.9395	0.8532	0.9159	0.9039

[‡] U-Net(128) / ResNet(101) for tissue seg. / patho. class.
[§] E2E-VarNet + U-Net(64) / ResNet(18) for tissue seg. / patho. class.

¶ see Supplement I

TABLE VII
COMPARISON OF AVERAGE TEST PSNR (dB) BETWEEN RECONSTRUCTION
MODELS TRAINED WITH TASK-SPECIFIC MASKS AND TACKLE_{recon} ON THE
FASTMRI KNEE DATASET.

Method	Brain seg.		Knee seg.		Tumor class.	
	16 $\times$	64 $\times$	16 $\times$	64 $\times$	16 $\times$	64 $\times$
Task-specific mask+VN	38.47	33.04	32.53	30.10	44.20	37.07
TACKLE _{recon} .	38.44	33.13	32.63	30.24	44.48	37.26

- Single larger predictor (row 1)
 - *Tissue seg.*: U-Net with 128 channels after the first convolution layer and the same number of pooling layers (42.2 M parameters)
 - *Patho. class.*: ResNet101 (42.5 M parameters)
- VN+predictor (rows 2&3)
 - *Tissue seg.*: E2E-VarNet + standard U-Net (29.9 M + 10.6 M = 40.5 M parameters)
 - *Patho. class.*: E2E-VarNet + ResNet18 (29.9 M + 11.2 M = 41.1 M parameters)

Comparing the first two rows, we find that the proposed “VN+predictor” architecture significantly outperforms the “single larger predictor” baseline on all settings. This is likely due to the model-based nature of the “VN+predictor” architecture, which more effectively extracts useful information from subsampled measurements for downstream tasks. Finally, we include the pre-training step discussed in Section III-D. Highlighted in cyan, the full-fledged version of TACKLE in the last row significantly outperforms the ablated baselines on both non-reconstruction tasks, indicating the importance of both the proposed architecture and training procedure.

C. Using Task-Specific Sequences for Reconstruction

Our optimized task-specific pipeline learns to adjust the image representation from a conventional form to one that is more readily interpretable by the predictor network. This often adds additional textures to the images, making them look different from traditional reconstructions. However, this does not imply there is a significant loss in information that could be used for image reconstruction. Despite being optimized for task-specific objectives, our learned task-specific subsampling patterns can be used retrospectively for generating high-fidelity reconstructions. To show this, we conduct an experiment where we take the

learned subsampling patterns of TACKLE_{seg} and TACKLE_{class} and train an additional reconstruction network for each subsampling pattern. The subsampling pattern is fixed during the training. This experiment mimics the scenario if one wants a traditional reconstruction out of the collected k -space samples from our task-specific sequences. In Table VII, we provide a comparison with TACKLE_{recon}, which jointly optimizes the subsampling pattern and reconstructor. One can see that the reconstruction models trained with task-specific masks (row 1) come close to TACKLE_{recon} (row 2) in terms of reconstruction performance. These results indicate that our task-specific models do not incur a significant loss of image information but achieve a better trade-off for the downstream task accuracy. It is thus possible to recover better images retrospectively using the k -space measurements collected by the task-specific sequences.

VII. LIMITATIONS

Building on the promising results we have achieved, we acknowledge opportunities for further improvement of our current study.

a) Data usage: Similar to other works on task-specific CS-MRI co-design, our approach requires matched k -space, image, and annotation labels, which are of limited quantity in the research community. Due to this limitation, two of our experiments (brain segmentation and tumor classification tasks) are conducted with k -space data simulated from magnitude images.

b) Sequence implementation: Although we have implemented a prospectively subsampling sequence with a learned sampling pattern by TACKLE_{ROI} on a Siemens MRI scanner, it was done using only one type of 3D gradient echo sequence. Other physical constraints affect the deployment of our method for general MRI sequences. For example, in spin-echo sequences, the order of sampling should be considered to mitigate spin-relaxation effects.

c) *Controlled study*: The evaluation in the current study is based on conventional quantitative metrics and qualitative visual comparisons. The number of volunteers for testing our learned sequences on a Siemens MRI scanner is also small. To further assess prospective subsampling, future evaluations should involve controlled studies of image quality with radiologists.

VIII. CONCLUSION

In this work, we generalized the objective of CS-MRI co-design to a variety of tasks beyond full-FOV reconstruction. We introduced TACKLE as a unified approach for robustly learning task-specific strategies. Through comprehensive experiments, we showed that TACKLE outperforms existing DL techniques that separately learn subsampling pattern, reconstruction, and prediction. Additionally, TACKLE outperforms naïve approaches to co-design that directly learn a direct mapping from measurements to predictions. We found that the optimized strategies sometimes circumvent the typical reconstruction in terms of point-wise accuracy, but effectively extract key visual information useful for task prediction. Through ablation studies, we justified multiple design choices with regard to architecture and training procedure, and showed their importance in effectively learning CS-MRI strategies for tasks that go beyond full-FOV reconstruction. We further implemented a learned subsampling sequence and tested it on a Siemens 3T MRI Skyra scanner, which led to a four-fold scan time reduction without sacrificing visual quality. Our study demonstrates the exciting promise of employing end-to-end co-design techniques, suggesting a future where clinical CS-MRI requirements are addressed with enhanced efficiency while maintaining accuracy.

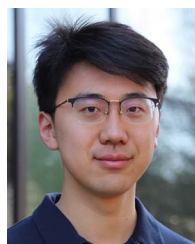
ACKNOWLEDGMENT

The authors would like to thank Xinyi Wu for her assistance as a volunteer in testing our learned MRI sequence and collecting data.

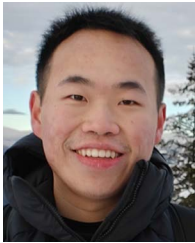
REFERENCES

- [1] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med., Official J. Soc. Magn. Reson. Med. / Soc. Magn. Reson. Med.*, vol. 58, pp. 1182–1195, 2007.
- [2] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.
- [3] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
- [4] R. Ahmad et al., "Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 105–116, Jan. 2020.
- [5] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "RARE: Image reconstruction using deep priors learned without groundtruth," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1088–1099, Oct. 2020.
- [6] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Int. Conf. 29th Adv. Neural Inf. Process. Syst., Annu. Conf. Neural Inf. Process. Syst.*, 2016, pp. 10–18.
- [7] 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. Imag.*, vol. 37, no. 2, pp. 491–503, Feb. 2018.
- [8] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1828–1837.
- [9] C. D. Bahadir, A. V. Dalca, and M. R. Sabuncu, "Learning-based optimization of the under-sampling pattern in MRI," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2019, pp. 780–792.
- [10] Z. Zhang, A. Romero, M. J. Muckley, P. Vincent, L. Yang, and M. Drozdal, "Reducing uncertainty in undersampled MRI reconstruction with active acquisition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2049–2058.
- [11] J. Zhang et al., "Extending LOUPE for k-space under-sampling pattern optimization in multi-coil MRI," in *Proc. 3rd Int. Workshop Mach. Learn. Med. Image Reconstruction*, 2020, vol. 12450, pp. 91–101.
- [12] H. K. Aggarwal and M. Jacob, "J-ModL: Joint model-based deep learning for optimized sampling and reconstruction," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1151–1162, Oct. 2020.
- [13] C. Alkan, M. Mardani, S. Vasanawala, and J. M. Pauly, "Learning to sample MRI via variational information maximization," in *Proc. NeurIPS Workshop Deep Learn. Inverse Problems*, 2020. [Online]. Available: https://openreview.net/forum?id=1AOReNkDmh_
- [14] G. Wang, T. Luo, J.-F. Nielsen, D. C. Noll, and J. A. Fessler, "B-Spline parameterized joint optimization of reconstruction and k-space trajectories (BJORK) for accelerated 2D MRI," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2318–2330, Sep. 2022.
- [15] S. Xue et al., "2D probabilistic undersampling pattern optimization for MR image reconstruction," *Med. Image Anal.*, vol. 77, 2022, Art. no. 102346. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841521003911>
- [16] W. Peng, L. Feng, G. Zhao, and F. Liu, "Learning optimal k-space acquisition and reconstruction using physics-informed neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 20794–20803.
- [17] G. Wang, J.-F. Nielsen, J. A. Fessler, and D. C. Noll, "Stochastic optimization of 3D non-cartesian sampling trajectory (SNOPY)," 2022, *arXiv:2209.11030*.
- [18] J. Yang et al., "Fast multi-contrast MRI acquisition by optimal sampling of information complementary to pre-acquired MRI contrast," *IEEE Trans. Med. Imag.*, vol. 42, no. 5, pp. 1363–1373, May 2023.
- [19] F. Martinini, M. Mangia, A. Marchioni, R. Rovatti, and G. Setti, "A deep learning method for optimal undersampling patterns and image recovery for MRI exploiting losses and projections," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 4, pp. 713–724, Jun. 2022.
- [20] J. Wang, Q. Yang, Q. Yang, L. Xu, C. Cai, and S. Cai, "Joint optimization of Cartesian sampling patterns and reconstruction for single-contrast and multi-contrast fast magnetic resonance imaging," *Comput. Methods Programs Biomed.*, vol. 226, 2022, Art. no. 107150. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169260722005314>
- [21] Z. Wang et al., "Leaders: Learnable deep radial subsampling for MRI reconstruction," in *Proc. IEEE 19th Int. Symp. Biomed. Imag.*, 2022, pp. 1–5.
- [22] G. Chaithya, Z. Ramzi, and P. Ciuciu, "Hybrid learning of non-cartesian k-space trajectory and MR image reconstruction networks," in *Proc. IEEE 19th Int. Symp. Biomed. Imag.*, 2022, pp. 1–5.
- [23] M. V. W. Zibetti, F. Knoll, and R. R. Regatte, "Alternating learning approach for variational networks and undersampling pattern in parallel MRI applications," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 449–461, 2022.
- [24] B. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.
- [25] E. J. Candès, J. K. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [26] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.
- [27] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [28] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D, Nonlinear Phenomena*, vol. 60, pp. 259–268, Nov. 1992.
- [29] H. Wang, D. Liang, and L. Ying, "Pseudo 2D random sampling for compressed sensing MRI," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2009, pp. 2672–2675.
- [30] S. Vasanawala et al., "Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, 2011, pp. 1039–1043.

- [31] N. Chauffert, P. Ciuciu, J. Kahn, and P. Weiss, "Variable density sampling with continuous trajectories," *SIAM J. Imag. Sci.*, vol. 7, no. 4, pp. 1962–1992, 2014.
- [32] R. L. Cook, "Stochastic sampling in computer graphics," *ACM Trans. Graph.*, vol. 5, no. 1, pp. 51–72, Jan. 1986, doi: [10.1145/7529.8927](https://doi.org/10.1145/7529.8927).
- [33] R. Raja and N. Sinha, "Adaptive k-space sampling design for edge-enhanced DCE-MRI using compressed sensing," *Magn. Reson. Imag.*, vol. 32, no. 7, pp. 899–912, 2014.
- [34] S. Venkatakrishnan, C. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 945–948.
- [35] U. Kamilov, C. Bouman, G. Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 40, pp. 85–97, Jan. 2023, doi: [10.1109/MSP.2022.3199595](https://doi.org/10.1109/MSP.2022.3199595).
- [36] S. Wang et al., "Accelerating magnetic resonance imaging via deep learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag.*, 2016, pp. 514–517.
- [37] D. Lee, J. J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing MRI," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, 2017, pp. 15–18.
- [38] D. Lee, J. Yoo, S. Tak, and J. C. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1985–1995, Sep. 2018.
- [39] G. Yang et al., "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.
- [40] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction with cyclic loss in generative adversarial networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1488–1497, Jun. 2018.
- [41] K. Hammernik et al., "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [42] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.
- [43] A. Sriram et al., "End-to-end variational networks for accelerated MRI reconstruction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2020, pp. 64–73.
- [44] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya, "Dense recurrent neural networks for accelerated MRI: History-cognizant unrolling of optimization algorithms," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1280–1291, Oct. 2020.
- [45] J. Liu, Y. Sun, W. Gan, X. Xu, B. Wohlberg, and U. S. Kamilov, "SGD-Net: Efficient model-based deep learning with theoretical guarantees," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 598–610, 2021, doi: [10.1109/TCI.2021.3085534](https://doi.org/10.1109/TCI.2021.3085534).
- [46] J. Adler and O. Ozan, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.
- [47] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magn. Reson. Med.*, vol. 84, no. 6, pp. 3172–3191, 2020.
- [48] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1123–1133, 2021.
- [49] J. Zbontar et al., "fastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*.
- [50] L. Zhao et al., "JoJoNet: Joint-contrast and joint-sampling-and-reconstruction network for multi-contrast MRI," 2022, *arXiv:2210.12548*.
- [51] J. Zou and Y. Cao, "Joint optimization of k-t sampling pattern and reconstruction of DCE MRI for pharmacokinetic parameter estimation," *IEEE Trans. Med. Imag.*, vol. 41, no. 11, pp. 3320–3331, Nov. 2022.
- [52] X. Xu, W. Gan, S. Kothapalli, D. Yablonskiy, and U. Kamilov, "CoRECT: A deep unfolding framework for motion-corrected quantitative R2* mapping," 2022, *arXiv:2210.06330*.
- [53] Z. Fan, L. Sun, X. Ding, Y. Huang, C. Cai, and J. Paisley, "A segmentation-aware deep fusion network for compressed sensing MRI," in *Proc. 15th Eur. Conf. Comput. Vis.*, 2018, pp. 55–70.
- [54] L. Sun, Z. Fan, Y. Huang, X. Ding, and J. W. Paisley, "Joint CS-MRI reconstruction and segmentation with a unified deep network," in *Proc. 26th Int. Conf. Inf. Process. Med. Imag.*, 2019, pp. 492–504.
- [55] T. Weiss et al., "PILOT: Physics-informed learned optimized trajectories for accelerated MRI," *Mach. Learn. Biomed. Imag.*, 2021.
- [56] Z. Wang et al., "One network to solve them all: A sequential multi-task joint learning network framework for MR imaging pipeline," in *Proc. 4th Int. Workshop Mach. Learn. Med. Image Reconstruction*, 2021, pp. 76–85, doi: [10.1007/978-3-030-88552-6_8](https://doi.org/10.1007/978-3-030-88552-6_8).
- [57] Z. Wang et al., "Promoting fast mr imaging pipeline by full-stack AI," *iScience*, vol. 27, no. 1, 2024, Art. no. 108608. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2589004223026858>
- [58] T. Yin et al., "End-to-end sequential sampling and reconstruction for MRI," in *Proc. Int. Conf. Mach. Learn. Health*, 2021, vol. 158, pp. 261–281. [Online]. Available: <https://proceedings.mlr.press/v158/yin21a.html>
- [59] Y. Bengio, N. Léonard, and A. Courville, "Estimating or propagating gradients through stochastic neurons for conditional computation," 2013, *arXiv:1308.3432*.
- [60] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.
- [61] T. Heimann, B. J. Morrison, M. A. Styner, M. Niethammer, and S. Warfield, "Segmentation of knee images: A grand challenge," in *Proc. MICCAI Workshop Med. Image Anal. Clinic*, 2010, pp. 207–214.
- [62] F. Schick, "Tissue segmentation: A crucial tool for quantitative MRI and visualization of anatomical structures," *Magn. Reson. Mater. Phys., Biol., Med.*, vol. 29, pp. 89–93, 2016.
- [63] A. Raj, S. Vishwanathan, B. Ajani, K. Krishnan, and H. Agarwal, "Automatic knee cartilage segmentation using fully volumetric convolutional neural networks for evaluation of osteoarthritis," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 851–854.
- [64] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, pp. 774–781, 2012.
- [65] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.
- [66] V. Ghodrati et al., "MR image reconstruction using deep learning: Evaluation of network structure and loss functions," *Quantitative Imag. Med. Surg.*, vol. 9, no. 9, 2019, Art. no. 1516.
- [67] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945. [Online]. Available: <http://www.jstor.org/stable/1932409>
- [68] K. H. Zou et al., "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, Feb. 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/14974593>
- [69] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. IEEE 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.
- [70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [71] R. Zhao et al., "fastMRI, clinical pathology annotations for knee and brain fully sampled magnetic resonance imaging data," *Sci. Data*, vol. 9, no. 1, 2022, Art. no. 152, doi: [10.1038/s41597-022-01255-z](https://doi.org/10.1038/s41597-022-01255-z).
- [72] A. D. Desai et al., "SKM-TEA: A dataset for accelerated MRI reconstruction with dense image labels for quantitative clinical evaluation," in *Proc. 35th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track (Round 2)*, 2021. [Online]. Available: https://openreview.net/forum?id=YDMFgD_qJuA
- [73] F. Knoll et al., "fastMRI: A publicly available raw k-space and dicom dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiol., Artif. Intell.*, vol. 2, no. 1, 2020, Art. no. e190007, doi: [10.1148/ryai.2020190007](https://doi.org/10.1148/ryai.2020190007).



Zihui Wu (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from the Washington University in St. Louis (WUSTL), St. Louis, MO, USA, in 2020. He is currently working toward the Ph.D. degree with the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA. His research interests include computational imaging, inverse problems, machine learning, and optimization. He is a Kortschak Scholar and the recipient of the Amazon AI4Science fellowship.



Tianwei Yin received the B.Sc. degree in computer science and mathematics from the University of Texas at Austin (UT Austin), Austin, TX, USA, in 2022. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. His research focuses on generative models and their applications in computer vision and graphics.



Andre van der Kouwe (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in electrical and electronic engineering from the University of Pretoria, Pretoria, South Africa, in 1992 and 1995, respectively, and the Ph.D. degree in biomedical engineering from the Ohio State University, Columbus, OH, USA, in 1999. He is currently an Associate Professor of radiology with the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital (MGH) and Harvard Medical School, Boston, MA, USA. His research focuses on MRI pulse sequence design and image analysis. He has been the key developer of sequences that are optimal with respect to brain morphometry, and structural sequences with embedded real-time motion correction that promise to open up structural imaging to an array of clinical populations that were difficult or impossible to image previously.



Yu Sun (Member, IEEE) received the B.Eng. degree in electronics and information from Sichuan University, Sichuan, China, in 2015, and the Ph.D. degree in computer science from the Washington University in St. Louis, St. Louis, MO, USA, in 2022. He is currently a Postdoctoral Fellow with the Department of Computational and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA. He will join the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA, as an Assistant Professor starting August 2024. He received the “Honor” evaluation by the Department of Computer Science, Washington University in St. Louis, from 2019 to 2021. His Ph.D. thesis was the winner of the Turner Dissertation Award in 2022. He is on the IEEE Signal Processing Society’s Computational Imaging Technical Committee.



Adrian V. Dalca received the B.Sc. degree in computer science from the University of Toronto, Toronto, ON, Canada, in 2008, and the Ph.D. degree in electrical engineering and computer science (EECS) from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2016. He is currently an Assistant Professor of radiology with the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, and a Research Scientist with the Computer Science and Artificial Intelligence Lab (CSAIL), EECS, MIT. His research focuses on machine learning techniques and probabilistic models for medical image analysis. Driven by scientific and clinical questions, he develops general and robust machine learning methods, solving tasks like registration and segmentation, that are broadly applicable and easy to use.



Robert Frost received the M.Phys. undergraduate degree in physics and D.Phil. degree in MRI physics from the University of Oxford, Oxford, U.K., in 2008 and 2013, respectively. He is currently an Assistant Professor of radiology with the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. He works on improving the quality and efficiency of brain MRI through modification of the acquisition and image reconstruction. His research focuses on techniques to measure head motion and associated MR effects during scans and thereby correct resulting artifacts in clinical and research MR neuroimaging. He was the recipient of the R21 Trailblazer Award from the National Institute of Biomedical Imaging and Biomedical Engineering (NIBIB) in 2021.



Katherine L. Bouman received the B.Sc. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2011, and the Ph.D. degree in electrical engineering and computer science (EECS) from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2017. She was a Postdoctoral Fellow with the Harvard-Smithsonian Center for Astrophysics, Cambridge. She is currently an Associate Professor with the Computing and Mathematical Sciences, Electrical Engineering, and Astronomy Departments, California Institute of Technology, Pasadena, CA, USA. Her work combines ideas from signal processing, computer vision, machine learning, and physics to find and exploit hidden signals for scientific discovery. As part of the Event Horizon Telescope Collaboration, she Co-Led the Imaging Working Group and acted as Coordinator for papers concerning the first imaging of the M87* and Sagittarius A* black holes. She is a Rosenberg Scholar, Heritage Medical Research Institute Investigator, recipient of the Royal Photographic Society Progress Medal, Electronic Imaging Scientist of the Year Award, Sloan Fellowship, University of Michigan Outstanding Recent Alumni Award, and co-recipient of the Breakthrough Prize in Fundamental Physics.