Graphical Abstract

Generalizing Supervised Deep Learning MRI Reconstruction to Multiple and Unseen Contrasts using Meta-Learning Hypernetworks

Sriprabha Ramanarayanan, Arun Palla, Keerthi Ram, Mohanasankar Sivaprakasam

Highlights

Generalizing Supervised Deep Learning MRI Reconstruction to Multiple and Unseen Contrasts using Meta-Learning Hypernetworks

Sriprabha Ramanarayanan, Arun Palla, Keerthi Ram, Mohanasankar Sivaprakasam

- A multi-modal meta-learning model for image reconstruction that provides mode-specific inductive bias closer to the target data distribution with deviated acquisition settings.

- Provides evolutionary capabilities of hypernetworks to dynamically predict weights that modulate the base reconstruction network weights for multimodal image reconstruction.

- Gradient-based meta-learning to optimize the kernel modulation hypernetworks in the contextual space.

- Extensive experimentation for multi-contrast MRI reconstruction to showcase the superior adaptation capabilities on-the-fly and via fine-tuning in a few gradient steps to unseen multi-contrast datasets.

- Superior reconstruction performance over joint training, other meta-learning methods, and various context-specific MRI reconstruction networks.

- Representational analysis showing that kernel modulation induces maximum mode-specific features in the high-resolution layers of the encoder-decoder base network.

Generalizing Supervised Deep Learning MRI Reconstruction to Multiple and Unseen Contrasts using Meta-Learning Hypernetworks

Sriprabha Ramanarayanan[a,b], Arun Palla[a,b], Keerthi Ram[b], Mohanasankar Sivaprakasam[a,b]

[a]*Department of Electrical Engineering, Indian Institute of Technology Madras (IITM), India*
[b]*Healthcare Technology Innovation Centre, IITM, India*

Abstract

Meta-learning has recently been an emerging data-efficient learning technique for various medical imaging operations and has helped advance contemporary deep learning models. Furthermore, meta-learning enhances the knowledge generalization of the imaging tasks by learning both shared and discriminative weights for various configurations of imaging tasks during training. However, existing meta-learning models attempt to learn a single set of weight initializations of a neural network that might be fundamentally restrictive under the heterogeneous (multimodal) data scenario. This work aims to develop a multimodal meta-learning model for image reconstruction, which augments meta-learning with evolutionary capabilities to encompass diverse acquisition settings of heterogeneous data. Our proposed model called KM-MAML (**K**ernel **M**odulation-based **M**ultimodal **M**eta-**L**earning), has hypernetworks (auxiliary learners) that evolve to generate mode-specific (or context-specific) weights. These weights provide the mode-specific inductive bias for multiple modes by re-calibrating each kernel of the base network for image reconstruction via a low-rank kernel modulation operation. Furthermore, we incorporate gradient-based meta-learning (GBML) in the contextual space to update the weights of the hypernetworks based on different modes. The hypernetworks and the base reconstruction network in the GBML setting provide discriminative mode-specific features and low-level image features, respectively. We extensively evaluate our model for multi-contrast magnetic resonance image reconstruction considering the essential research directions in fastMRI for multimodal and rich transfer learning capabilities across var-

ious MRI contrasts. Our comparative studies show that the proposed model (i) exhibits superior reconstruction performance over joint training, other meta-learning methods, and various context-specific MRI reconstruction architectures, and (ii) better adaptation to 80% and 92% of unseen multi-contrast data contexts with improvement margins of 0.1 to 0.5 dB in PSNR and around 0.01 in SSIM, respectively. Besides, a representation analysis with U-Net as the base network shows that kernel modulation infuses 80% of mode-specific representation changes in the high-resolution layers. Our source code is available at https://github.com/sriprabhar/KM-MAML/.

1. Introduction

Deep learning methods have seen remarkable improvements in various imaging tasks like image classification, recognition, and reconstruction. The success of these methods has vastly been in scenarios where the model is trained and tested on a homogeneous dataset representing a specific concept like similar objects, characteristic features, image structures, and contrast levels. The homogeneous conditions across training and test data limit the robustness of the model when the image data deviates due to shifts in scan settings and contrasts [1]. For instance, in diverse medical imaging systems like magnetic resonance imaging (MRI), devising robust models that generalize to multi-scanner data is crucial for transferring these models into clinical practice [2]. Recent advances aim to improve the robustness of the model by exploring in two promising directions [3], (i) At the data level, obtaining heterogeneous modalities (multimodal image data) with multiple imaging settings to learn from diverse data (ii) At the model level, incorporating adaptive learning mechanisms to improve the generality of the model to varying imaging conditions. To handle challenging scenarios with multi-modal data distributions encompassing diverse acquisition settings, an adaptive learning model that can provide the shared knowledge and discriminative features of varied modalities is required [4, 5, 6].

Multimodal image data refers to the outputs of image acquisition devices (like sensors or acquisition technologies) with multiple intensity representations of visual concepts [3] (Figure 1(left)). Integrating multiple perspectives of concepts using multimodal data enhances the knowledge generalization

2

Figure 1: Graphical abstract of MRI data sources and learning methods (best viewed in color). Left: The multimodal MRI data sources (datasets), each represented by various shapes and colors. Each row corresponds to a specific mode of the multimodal MRI data and covers a specific contrast, with different acquisition settings like different types and amounts of under-sampling degradation. Right: Joint training has a single level of optimization of the base network by combining the training samples from the datasets. Meta-learning has inner (gray arrow) and outer (green arrow) levels of optimization. We categorize each dataset corresponding to each contrast and scan setting as a training task for the neural network. Each task consists of support and query partitions of the dataset. The inner and outer levels of optimization use the support and query partitions of each task data.

of the concept by learning a shared representation across modalities, eliminating models trained independently for each modality [7, 8]. The learning approach commonly adopted to integrate multimodal data is joint training using stochastic gradient descent (SGD) (Figure 1 (right)), Joint training). However, due to statistical shifts across various modes [7], the shared knowledge gained by joint training from image samples alone might be inadequate and could underfit modalities with deviations in the acquisitions settings from the training data [7, 4]. Secondly, the learning process does not consider discriminative features, meaning that it lacks an adaptive mechanism of learning unique sets of weights for each modality and encapsulating relationships across modalities in a common weight space [6].

Recently, meta-learning methods [9] have emerged as a data-efficient

Figure 2: Concept diagram of the bi-level optimization of MAML and KM-MAML (best viewed in color). Left: MAML involves bi-level optimization (gray arrows for the inner loop and green arrows for the outer loop) of the base learner (blue box). Different from MAML, KM-MAML involves two networks - a base learner for the imaging task and an auxiliary network called the kernel modulation (KM) network to infer mode-specific weights. The two levels of optimization are 1) outer loop or meta-weight updates of the base network θ and the KM network ω shown as green arrows backward. 2) The inner loop or task-specific adaptation via a few gradient updates of the KM network (gray arrows within the pink box). The meta-parameters of the KM network infuse mode-specific knowledge to the base learner via kernel modulation to create an improved base learner. (Right) Weight update process in the task space. KM-MAML provides mode-specific initializations (red arrows pointing to multi-colored hexagons) that coarsely capture the target MRI data distribution, unlike MAML, which has a single meta-initialization. Fine-tuning the base network or the hypernetwork by a few gradient steps (gray arrows) refines the model closer to the target data distribution (Algorithm 1 shows the training details).

adaptive learning technique for regression and classification methods, both at the task (a unique mode) and the sample level. Model-agnostic meta-learning (MAML) [10] learns a common prior knowledge across tasks and the task-specific parameters of a network in a bi-level interleaved optimization process (Figure 1 (right bottom)), enabling faster adaptation to new tasks using the learned prior (Figure 2, MAML). MAML outperforms conventional learning [11]; however, when the data distribution of tasks is multimodal, the knowledge gained by MAML is fundamentally restricted by a single initialization that is learned by a single (base) network [12] performing an imaging operation. For instance, it would be infeasible to seek a common initialization for an imaging task with continuously varying acquisition settings.

Within the premises of heterogeneous data, the data samples occupy different regions in the same high dimensional space as clusters that correspond to the modes of the data. However, the clusters might not lie closer to each other as the degree of similarity between them might be vastly different. Due to variations in the acquisition settings, different modes are endowed with complementary properties that might contribute differently to the learning process. This motivates the need for adaptive context-aware (or mode-specific) initializations to provide reliable and selective meta-knowledge closer to the target data distribution with deviated acquisition settings. We consider the problem of image restoration using a single model that can scale to multiple acquisition settings of multi-modal data by generating mode-specific initializations for rapidly adapting to unseen data.

Task-aware modulation of the model weights using Multi-modal meta-learning (MMAML) [13] is a promising direction that infers mode-specific latent representations to capture the features corresponding to the modes of heterogeneous data distribution. Our method uses auxiliary networks to execute mode-specific modulation of the weights of a base convolutional neural network (CNN) using the MMAML approach, in contrast to adaptive strategies that adjust the activations of the base CNN. In order to improve upon MMAML, our focus lies on how optimally the two networks learn together with two objectives. The objective of the auxiliary networks is to provide the high-level inductive bias [10] of multiple modes of heterogeneous data, and the objective of the base network is to solve the imaging tasks. Unlike previous MMAML methods [13, 12], which infuse adaptive mechanism only on the base network via meta-learning, our approach infuses adaptive learning on both the networks at the model and optimization levels to achieve the two objectives.

(i) *Model level:* Inspired by the evolutionary computing [14, 15] capabilities of weight-learning networks [16, 8, 6], we propose adaptive hypernetworks [17], called kernel modulation (KM) hypernetworks, consisting of a low-rank KM layer to provide mode-specific inductive bias to the base CNN. The human brain exhibits attentional modulation, wherein modulatory signals representing behavioral goals evolving in brain regions are sent to sensory neurons in the cortex for perception and contextual modulation [18]. Similarly, the proposed KM hypernetworks evolve to generate different weights to modulate kernels in the base network for dynamically re-calibrating the behavior of the base network.

(ii) *Optimization level:* We categorize the hypernetworks' weights as context-based, and the base network's weights as image feature-based. The KM hypernetworks are optimized using bi-level gradient-based meta-learning inspired by the learning-to-learn [16] approach in the contextual space. The base network is optimized via both gradient updates, and mode-specific kernel modulation parameters from hypernetworks [19] (Figure 2 KM-MAML). At test time, the base network is adapted on-the-fly or fine-tuned using KM for multi-modal imaging tasks.

We demonstrate the efficacy of our approach in multi-contrast magnetic resonance image (MC-MRI) reconstruction [20]. MC-MRI captures diverse and complementary perspectives of a single subject using multiple MRI contrasts, each representing a mode. As charted out in the fastMRI reconstruction challenge results [21], the pre-existing MC-MRI deep learning models [22, 23, 24] lack scalability to different contrasts and exhibit inadequate transfer capabilities across MRI modalities. We take a kernel modulation-based meta-learning approach to identify the modes of each MRI contrast and scale to multiple contrast-specific models. We summarize our contributions as,

1. We propose a meta-learning model called KM-MAML, that learns via context-aware kernel modulation for multimodal image reconstruction. The proposed model has a base reconstruction network and layer-wise hypernetworks, which dynamically modulate each kernel of a base reconstruction network based on the mode of the multimodal data.

2. The proposed model is optimized with two training objectives, extracting (i) the mode-specific inductive bias from hypernetworks by adopting learning-to-learn in the contextual space of heterogeneous data and (ii) low-level image features from the base network.

3. Extensive experimentation on MC-MRI reconstruction shows that the

proposed model provides superior reconstruction performance over other learning methods, generality to 80% and 92% of unseen multi-contrast contextual settings in PSNR and SSIM, with improvement margins of around 0.1 dB in PSNR and 0.01 in SSIM, respectively. Our approach also matches the performance of various context-specific MRI reconstruction architectures.

4. Our analysis of the representational similarity between pre and post kernel modulation features shows that with U-Net as the base network, kernel modulation induces maximum (of 80%) discriminative mode-specific representations in the top (highest resolution) layers of the encoder and decoder.

This paper is organized as follows. Section 2 describes the related works. Section 3 provides the material and methods with algorithm and architecture details. Section 4 and 5 provides the dataset, implementation details, and results. Section 6 briefly summarizes our findings and conclusion.

2. Related Work

The concept of meta-learning is originally developed by the learning-to-learn [16, 9, 25] approach which is further extended using evolutionary computing (EC) [26, 27, 28, 15] methods to learn the rules. The emerging directions in multimodal learning have shown that the training data acquired from one modality can benefit from the shared knowledge across different modalities [29, 30]. The focus of multimodal meta-learning is to effectively learn a mode-specific prior using cross-modal discriminative features to adapt to several unseen multimodal data [3].

2.1. Evolutionary computing and Meta-learning

Recent works that relate evolutionary computing and deep learning, include differentiable compositional pattern producing network (DPPN) [31], hypernetworks [17, 32], and population-based meta-learning (PBML) [33]. DPPN and hypernetworks use a multi-layer perceptron (MLP) network to directly evolve the structure and weights of another neural network. PBML relates EC with optimization-based meta-learning (MAML) by sharing the paradigm of learning to learn. The authors of PBML show that MAML manifests adaptive evolvability [34] and provides inductive biases that balance exploration and exploitation along various dimensions of mutation functions. In this context, we motivate that ours is an evolutionary deep learning

7

model with weight learning capability and adaptive evolvability using hyper-networks and MAML, respectively, for image reconstruction.

2.2. Hypernetworks

Hypernetworks are mainly inspired by HyperNEAT [35], a neuroevolution framework wherein a network could be more intuitively evolved, node by node, and connection by connection to produce connectivity patterns. Initially developed for model compression [17], hypernetworks find various imaging applications like video frame prediction [36], point-cloud up-sampling [37], and multiple parameterized image restoration operators [6, 8]. These are model-based meta-learning methods [16, 38] wherein the hyper-networks predict all the parameters of the imaging task-oriented base network. In our work, the KM hypernetworks are not the sole source of base network weights. As a result, our model benefits from context-invariant knowledge learned by the base network. In interactive image restoration applications [39, 5, 40], hypernetworks are used to modulate the activations of the base convolution layer. In our model, the hypernetworks provide mode-specific inductive bias via kernel modulation for multi-modal imaging operations.

2.3. Gradient-based Meta-learning

Several variants of gradient-based meta-learning (GBML) or MAML [41, 42, 43, 44, 45] have shown promising results in few-shot learning, owing to its potential for rapid adaptation and feature reuse capabilities to future tasks [46]. MAML methods that focus on improving inner loop optimization using a part of the network architecture for classification and regression tasks exist in the literature. These methods, namely MeTAL [47], HyperMAML [19], and ALFA [48] use hypernetworks to learn task adaptive loss function, task-specific gradient updates, and inner loop regularization hyperparameters respectively. The meta-modulated CNN for snapshot compressive sensing (Meta-SCI) [49] uses a set of parameters that are used for KM based on different mask settings during test-time adaptation. The conditional neural adaptive process (CNAP) [4] uses a linear classifier as an adaptation network for classification tasks. In context adaptation via meta-learning (CAVIA) [50], a set of contextual parameters are meta-learned for adaptation to multiple tasks.

Multi-modal meta-learning methods, task-aware modulation (MMAML) [13], and contrastive knowledge distillation meta-learning (CAML) [12] modulate the base network layer with a single mode-specific parameter. Unlike

8

these works, which focus on bi-level optimization of only the base network, our work efficiently utilizes both the networks based on task-specific weight updates via kernel modulation of the base network and bi-level gradient descent updates of our proposed KM hypernetwork. The KM layer outputs modulation weights based [49, 51] on each sample to modulate each kernel of the base network.

2.4. Hypernetworks for MRI reconstruction

Methods that use hypernetworks for MRI reconstruction include the Hyperrecon network [52, 53] and the MAC-ReconNet [8]. The MAC-ReconNet provides scalable reconstructions for multiple acquisition settings. The Hyperrecon network provides regularization-agnostic hyperparameters for image reconstruction. Unlike these methods, we use both the base network and the modulation networks with GBML to provide a good initialization for adapting to unseen tasks.

The hypernetwork-based methods aforementioned [8, 5] lack interaction with the base network weights to achieve mode-specific weight initializations. As introduced before, different MRI contrasts exhibit different intensity properties and contribute to the learned features differently. This inspires us to employ MMAML for enhancing the generalization ability of the base neural network using meta-learned hypernetworks. Existing MMAML-based methods exhibit a resemblance to vanilla MAML as the modulation network is optimized in the outer level of optimization, emphasizing context-invariant learning. Secondly, MMAML infers a single mode-specific value that globally modulates all the kernels of the base network layer. Thirdly, these methods focus on classification tasks and inherently assume the presence of ground truth labels at test time for fine-tuning. We address these limitations by employing 1) low-rank kernel modulation offering multiplicative interactions [54] by dynamically modulating each kernel of the base network with different heterogeneous context-specific weights 2) two-level optimization of the hypernetworks' weights to lay emphasis on contextual learning. Furthermore, to support fine-tuning and on-the-fly adaptation to unseen contexts, our method employs only the input images to generate the task embeddings for conditioning the hypernetworks.

3. Material and Methods

3.1. Problem Formulation

The data acquisition forward model of the MR image reconstruction problem [55] can be formulated as:

$$\mathcal{A}x + \epsilon = y \tag{1}$$

where, $x \in \mathbb{C}^N$ denotes the desired image, $y \in \mathbb{C}^{M'}$ is the under-sampled (US) measurement from the MRI scanner, $\epsilon \in \mathbb{C}^{M'}$ is the noise and $\mathcal{A} : \mathbb{C}^N \rightarrow \mathbb{C}^{M'}$ represents the forward operator of the MRI acquisition process that causes aliasing artifacts due to k-space under-sampling. The under-sampled image reconstruction is ill-posed because the problem is under-determined ($M' << N$) and the operator $\mathcal{A}$ is ill-conditioned. The under-sampled or zero-filled (ZF) image is given by, $x_{US} = F_{US}^H y$ where F_{US} is the under-sampled Fourier encoding matrix. The reconstruction of the under-sampled image is achieved by introducing an apriori knowledge of x into the unconstrained optimization [55] given as:

$$\min_{x} \quad ||\mathcal{A}x - y||_2^2 + \mathcal{R}(x) \tag{2}$$

where, $||\mathcal{A}x - y||_2^2$ is the data fidelity term [56] and $\mathcal{R}$ is a regularization term.

Deep learning-based MRI reconstruction involves training a deep learning (DL) model using a single-level optimization on the average loss of all observed data samples. This supervised joint training can be formulated as:

$$\theta^* = \underset{\theta}{\text{argmin}} \quad \underset{(x_{US}, x_{FS}) \in \bigcup_i \mathcal{D}_i}{\mathbb{E}} [||x_{FS} - f(x_{US}; \theta)||_2^2] \tag{3}$$

where, $\mathcal{D}_i$ represents the dataset of the ith configuration of MRI contrast, under-sampling mask type, and acceleration factor, consisting of ground truth or fully sampled (FS) image x_{FS} and its corresponding under-sampled image x_{US}. Here, f is the DL model parameterized by θ with k-space data fidelity. Unlike iterative methods in Eq. 2, the formulation in Eq. 3 infers an optimal parameter set θ^* [16].

In MAML-based MRI reconstruction, we consider each combination of MRI contrast, under-sampling mask type, and acceleration factor for under-sampling as a task. We partition the data of each task M into support

$(x_{US,spt}^M, x_{FS,spt})$ and query $(x_{US,qry}^M, x_{FS,qry})$ samples. The parameters of the network, θ, in Fig. 2 (MAML) are called *meta-initializations*. For every task M, MAML uses the support samples to perform a few gradient-descent steps (*adaptation*) from meta-initializations to obtain task-specific parameters ϕ_M (Eq. 5). The loss due to task-specific parameters on query data is aggregated over all train tasks $(\mathcal{T})$, to provide supervision for meta-initializations (Eq. 4). The MAML optimization [16] is given as:

$$\theta^* = \operatorname*{argmin}_{\theta} \sum_{M \in \mathcal{T}} ||x_{FS,qry} - f(x_{US,qry}^M; \phi_M)||_2^2 \tag{4}$$

$$s.t \quad \phi_M = \operatorname*{argmin}_{\theta} ||x_{FS,spt} - f(x_{US,spt}^M; \theta)||_2^2 \quad \forall M \in \mathcal{T} \tag{5}$$

In the proposed KM-MAML, the parameters of the KM hypernetworks are optimized in the bi-level optimization and the parameters of the base network are optimized in the outer loop as shown in Fig. 2 (KM-MAML). The KM-MAML formulation is given by:

$$\omega^*, \theta^* = \operatorname*{argmin}_{\omega, \, \theta} \sum_{M \in \mathcal{T}} ||x_{FS,qry} - f(x_{US,qry}^M; \Omega_M, \theta)||_2^2 \tag{6}$$

$$s.t \quad \Omega_M = \operatorname*{argmin}_{\omega} ||x_{FS,spt} - f(x_{US,spt}^M; \omega, \theta)||_2^2 \quad \forall M \in \mathcal{T} \tag{7}$$

The training process of KM-MAML is illustrated in Algorithm 1. On every task's support samples, KM-MAML performs adaptation of the KM hypernetworks to result in a task-specific KM hypernetwork model that is characterized by weights Ω_M (Eq. 7), and this constitutes the inner level optimization in the bi-level MAML process (steps 5 to 14 in Algorithm 1). Step 13 in Algorithm 1 describes the L_1 image reconstruction loss due to the task-specific parameters Ω_u^M at the uth inner step and the modulated base network weights θ_{mod} using support data D_{spt}^M of task M. The loss due to task-specific parameters on query data, $L([\Omega_{U-1}^M, \theta_{mod}], D_{qry}^M)$ is aggregated over all train tasks $(\mathcal{T})$ in step 20, to provide supervision for meta-initializations ω and θ (Eq. 6 and steps 22 and 23). The low-rank modulation weights α and β and the modulated base network weights θ_{mod} at the uth inner update are shown in steps 10 to 12 and for the outer loop, in steps 18 to 20 for each task ($\odot$ denotes element-wise multiplication). The outer loop or meta-update of the base network and the hypernetworks weights are shown in steps 22 and 23, respectively. These weights, θ and ω form the meta-initializations of the proposed network.

Algorithm 1 KM-MAML Training Algorithm

Require: Learning rates η_1, η_2 and $p(\mathcal{T})$: Multimodal train task distribution

Require: $CE(.)$: Context encoder

 1: Randomly initialize ω: modulation network weights and θ: base network weights

 2: **for** $e = 1$ *to max_epochs* **do**

 3:     Sample a task mini-batch: $\mathcal{T}_{batch} \sim p(\mathcal{T})$

 4:     $\mathcal{L}_{meta} = 0$

 5:     **for** each training task M in $\mathcal{T}_{batch}$ **do**

 6:         Sample a mini-batch of support data: $D_{spt}^M = \{x_{US,n}^M, x_{FS,n}\}_{n=1}^{N_{spt}}$

 7:         Initialize $\Omega_0^M \leftarrow \omega$

 8:         **for** $u = 0$ *to* $U - 1$ **do**

 9:             Context embedding: $\gamma = CE(x_{US}^M)$ $\triangleright$ Support data mini-batch input

10:             $\alpha, \beta = $ KM Hypernetwork$(\gamma; \Omega_u^M)$

11:             $W_M = \beta \otimes_{outer} \alpha$ $\triangleright$ Modulation weights of task M

12:             $\theta_{mod} = \theta \odot W_M$ $\triangleright$ Mode-specific initialization (KM)

13:             $\mathcal{L}_1 = L([\Omega_u^M, \theta_{mod}], D_{spt}^M)$

14:             $\Omega_{u+1}^M \leftarrow \Omega_u^M - \eta_2 \nabla_{\Omega_u^M} \mathcal{L}_1$ $\triangleright$ mode-specific weight updates

15:         **end for**

16:         Sample a mini-batch of query data: $D_{qry}^M = \{x_{US,n}^M, x_{FS,n}\}_{n=1}^{N_{qry}}$

17:         $\alpha, \beta = $ KM Hypernetwork(γ, Ω_{U-1}^M)

18:         $W_M = \beta \otimes_{outer} \alpha$ $\triangleright$ $\otimes_{outer}$ denotes outer product

19:         $\theta_{mod} = \theta \odot W_M$

20:         $\mathcal{L}_{meta} \leftarrow \mathcal{L}_{meta} + L([\Omega_{U-1}^M, \theta_{mod}], D_{qry}^M)$

21:     **end for**

22:     $\theta \leftarrow \theta - Adam[\eta_1, \mathcal{L}_{meta}]$

23:     $\omega \leftarrow \omega - Adam[\eta_1, \mathcal{L}_{meta}]$ $\triangleright$ Meta-updates

24: **end for**

3.2. Architecture

The architecture of KM-MAML consists of a context encoder, a base reconstruction network, and layer-wise KM hypernetworks as shown in Figure 3. The context encoder is a simple auto-encoder that learns to map the under-sampled MRI input image to output in an unsupervised manner. The latent layer of the context encoder has c channels which are averaged to obtain a latent c-dimensional embedding vector which is passed to the KM

hypernetworks. The input and output of the base reconstruction network are the under-sampled, and the predicted fully sampled MRI images, respectively. The base reconstruction network is an encoder-decoder network, U-Net [57, 58], a popular benchmark for MRI reconstruction [21]. The base network has seven layers with three sub-sampling levels of the encoder and decoder each and a bottleneck layer.

There are seven KM hypernetworks corresponding to each layer of the base network. Each KM hypernetwork is a linear multi-layer perceptron network with three layers - an input layer with c units, a bottleneck layer, and an output layer with a size equal to the number of filters in the corresponding layer of the base network. The KM hypernetwork of layer l gives two rectangular matrices as output, $\beta^l \in R^{n_{out}^l \times r}$ and $\alpha^l \in R^{r \times n_{in}^l}$ where n_{out}^l and n_{in}^l are the number of output and input channels of the base network layer l and r is the rank. The outer product operation between β^l and α^l is computed in the KM layer (shown in steps 11 and 18 in Algorithm 1) to obtain modulation weights, W_M^l given as,

$$W_M^l = \beta^l \otimes_{outer} \alpha^l \qquad (8)$$

The modulation weights are specific to each kernel of the base network and are element-wise multiplied. For instance, if the base network filter size is $64 \times 32 \times 3 \times 3$, then the corresponding KM hypernetwork has $64 + 32$ output neurons, 64 for β and 32 for α. The outer product at the output gives modulation weights W_M^l of size 64×32 for kernel modulation.

4. Experiments and Results

4.1. Datasets and Implementation Details

4.1.1. Datasets

We use three multi-contrast MRI datasets consisting of axial brain images, namely **MRBrainS** [59], **IXI**[1] and **SRI24 Atlas** [60, 22].

1) **MRBrainS** dataset: We consider 336 slices of T1 and FLAIR contrasts from 7 volumes of T1 and FLAIR (fluid-attenuated inversion recovery) with T1: TR = 7.9 ms, TE = 4.5 ms and FLAIR: TR = 11s and TE = 125 ms acquired on a Philips scanner. The repetition and echo times, TR, and

---

[1]https://brain-development.org/ixi-dataset/

Figure 3: Network architecture of KM-MAML: The context-encoder CNN provides embedding vectors to represent each mode. The KM hypernetworks constitute the evolutionary computing module which predicts layer-wise weights that modulate the base network weights via low-rank kernel modulation layer (orange block), shown with rank $r = 1$ in the top right. The base CNN performs the multimodal image reconstruction task. X_{CNN} denotes the predicted image of the base network.

TE denote the contrast-specific MRI settings. All the images have a size of 240×240.

2) **IXI** dataset: We consider 1400 slices of T2 and Proton density (PD) weighted contrasts acquired from 14 volumes with T2: TR = 5725 ms, TE = 100 ms, and PD: TR = 5725 ms, TE = 8 ms acquired on a Philips scanner. The images are preprocessed by cropping to a size of 240×240.

For training the models we consider T1, FLAIR and T2 and PD contrasts from **MRBrainS** and **IXI** datasets.

3) **SRI24 Atlas** dataset: 135 slices of T1, T2, and PD-weighted MRI images with T1: TR = 6.5 ms, TE = 1.54 ms and T2, PD: TR = 10s, TE = 14/98 ms acquired on a GE scanner. All the images have a size of 240×240.

For testing the models on unseen scenarios, we use T1, T2, and PD from **SRI24 Atlas** dataset.

For comparison with other MRI reconstruction architectures, we use the following datasets:

4) **Automated Cardiac Diagnosis Challenge (ACDC) [61] cardiac MRI** dataset which consists of 150 and 50 patient records with 1841 and 1076 slices for training and validation respectively. The 2D slices are extracted and cropped to 150×150. The masks for radial under-sampling are taken from RefineGAN[2] repository

5) The **knee dataset** contains 200 training and validation single channel slices of 10 subjects, each obtained from a 15-element coil knee data[62] cropped to 320×320.

6) For training on large-scale data, we use the **fastMRI** [21] dataset with 7040 training and 3317 validation slices of 200 and 99 volumes, respectively, of coronal proton-density without (PD) and with fat suppression (PDFS). We split the validation slices into 437 slices from 18 volumes for adaptation and 2880 slices for evaluation. The slices have multiple resolution levels - 640×368, 640×372, 640×400, 640×454, and mask patterns different from the training data.

4.1.2. Implementation Details

Training details: For training and evaluating the models, we use four MRI contrasts - T1, FLAIR, T2, and PD. In addition, we augment the datasets based on two types of under-sampling mask patterns - Cartesian and Gaussian masks with three different acceleration factors for under-sampling namely, 4x, 5x, and 8x. We pose each configuration of contrast, mask type, and acceleration factor as a task for meta-learning. The configurations above are combined to form 24 tasks. The US input images are obtained by retrospectively under-sampling the fully sampled k-space [63]. Each task M has samples split into support D_{spt}^{M} and query D_{qry}^{M} images for meta-training.

For evaluating the adaptability to unseen tasks (from SRI24 Atlas), we choose three contrasts - T1, T2, and PD, two mask types, and five accelerations - 5x, 6x, 7x, 8x, and 9x to obtain 30 tasks. For adaptation via fine-tuning, we split the test images into test support for fine-tuning and test query images for evaluation. Table 1 shows the details of the number of sam-

---

[2]https://github.com/tmquan/RefineGAN

Table 1: Details of cross-scanner datasets, and the number of samples used for training and testing the models. Note that for SGD the support and query samples are mixed.

Dataset type	Dataset / scanner name	Contrast	Number of data samples		
			Support	Query	Testing
Training datasets	MRBrainS / Philips	T1	3456	2304	2304
		FLAIR	3456	2304	2304
	IXI / Philips	PD	4800	4800	24000
		T2	4800	4800	24000
Unseen datasets	SRI24 Atlas / GE	T1	-	-	1350
		T2	-	-	1350
		PD	-	-	1350

ples used for training the models. All MAML models are trained with one inner loop gradient step. The batch size for support and query data (N_{spt}, N_{qry}) is 10. The task mini-batch size is chosen as 3. The loss function is L_1 norm between the model's prediction and ground truth. We have chosen the learning rates η_1, η_2 of 0.001 with 600 epochs for training. For SGD, the learning rate is chosen as 0.001. All models are implemented using PyTorch and trained on $Nvidia RTX - 3090$ GPU with 24 GB memory. We have chosen the context embedding vector size as 256 and the rank $r = 1$ for kernel modulation. Our source code is available at https://github.com/sriprabhar/KM-MAML/.

Evaluation metrics: Our evaluation metrics are Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), to assess the performance of the models, and Centered Kernel Alignment (CKA) [64] to quantify the % change in the mode-specific features pre and post kernel modulation.

K-space Data fidelity: The output of the base network is fed to a k-space data fidelity unit to ensure consistency with the acquired k-space [56] (Appendix A.1).

Architecture-based Comparison: We compare our method with image domain MRI reconstruction CNNs, DAGAN [55], DC-CNN [56], DC-DEN [65], DC-RDN [66], DC-UNet [58], MICCAN [67], and MAC-ReconNet [8]. We use the source code of DC-CNN[3], DAGAN[4] and MAC-ReconNet[5] for implementation. We have also compared our method with recent state-of-the-art methods that have been proposed for MRI reconstruction. These are OUCR (Over-complete and under-complete CNNs) [68], vision transformer-based methods, ViT [69], and SWIN transformer [70] models, and a knowl-

---

[3]https://github.com/js3611/Deep-MRI-Reconstruction
[4]https://github.com/tensorlayer/DAGAN
[5]https://github.com/sriprabhar/MAC-ReconNet

edge distillation-based MRI reconstruction method, SFT-KD-Recon[6] [71]. For DC-UNet, DC-RDN, MICCAN, DC-DEN, ViT and SWIN transformers, we replace the CNN in DC-CNN with UNet from the fastMRI repository[7], dilated convolutions with recursive connection, channel attention, dense connection blocks, vision transformers and SWIN transformers respectively [72].

4.2. Results and Discussion

We extensively evaluate our approach by comparing it with three representative models, joint training using SGD, MAML [10], and MMAML [13] for MC-MRI reconstruction. Our experiments showcase (i) the scalability of the models in MC-MRI reconstruction, (ii) generalization to unseen contrasts via adaptation on the fly and fine-tuning, (iii) the contribution of KM and the benefits of GBML using comparative studies and representational similarity measures, (iv) an ablative study with and without meta-learning, (v) performance comparison with several MRI reconstruction architectures for specific and multiple acquisition contexts, and (vi) fine-tuning to new anatomies and image resolution levels.

4.2.1. Multi-modal MRI reconstruction performance

We evaluate the scalability of the meta-initializations of the model to training tasks. Table 2 shows the quantitative results for 12 out of 24 tasks. From the table, our observations are. (i) MAML-based models enrich the learning process compared to joint training indicated by improved SSIM metrics. (ii) KM-MAML consistently performs better than other methods for 96% of tasks in SSIM and 80% in PSNR. (iii) MMAML performs on par with MAML due to its closer resemblance to MAML, constraining the meta-training process to only the base network. (iv) KM-MAML consistently performs better than MMAML, highlighting the importance of meta-learning the auxiliary network for efficient adaptation.

The box plots for each configuration of contrast, mask types, and acceleration factors (Figure 4(a - d)) and each MRI contrast separately (Figure 4 (e)) together show that KM-MAML can provide improved accuracy metrics for multiple modes. The reliability of KM-MAML is demonstrated with relatively lesser deviations in box plots as compared to other methods. The

---

[6]https://github.com/gayathrimatcha/sft-kd-recon
[7]https://github.com/facebookresearch/fastMRI

Table 2: Quantitative comparison of ZF, SGD, MAML, MMAML, and KM-MAML for multiple training tasks combining various contrasts, acceleration factors, and mask types. Green and blue colors indicate the best and the second best metrics respectively. The tasks are denoted in short as contrast type - mask type - acceleration factor (T1C8 indicates T1 contrast with Cartesian mask pattern and 8x acceleration). FL denotes FLAIR MRI

Task	ZF PSNR / SSIM	SGD PSNR / SSIM	MAML PSNR / SSIM	MMAML PSNR / SSIM	KM-MAML PSNR / SSIM
T1C5	31.38 / 0.665	35.03 / 0.866	35.22 / 0.875	35.14 / 0.876	35.58 / 0.895
T1G4	32.24 / 0.545	40.07 / 0.913	40.59 / 0.922	40.71 / 0.927	41.02 / 0.935
T1G8	29.17 / 0.451	35.03 / 0.805	35.33 / 0.820	35.51 / 0.841	35.96 / 0.859
FLC4	28.40 / 0.642	34.12 / 0.879	34.40 / 0.885	34.23 / 0.882	34.40 / 0.889
FLC8	26.49 / 0.589	30.83 / 0.813	31.02 / 0.828	30.89 / 0.824	31.20 / 0.841
FLG8	26.36 / 0.426	33.40 / 0.802	33.73 / 0.811	33.56 / 0.820	33.93 / 0.832
PDC4	25.68 / 0.625	32.76 / 0.877	33.06 / 0.882	32.96 / 0.879	33.24 / 0.888
PDC8	23.40 / 0.551	29.07 / 0.803	29.48 / 0.814	29.32 / 0.804	29.64 / 0.819
PDG4	26.88 / 0.541	37.78 / 0.908	37.98 / 0.911	37.81 / 0.908	38.04 / 0.914
T2C4	25.94 / 0.634	32.01 / 0.873	32.43 / 0.880	32.26 / 0.875	32.55 / 0.884
T2G4	27.16 / 0.543	37.27 / 0.910	37.48 / 0.912	37.21 / 0.906	37.59 / 0.916
T2G5	25.71 / 0.500	34.83 / 0.860	35.06 / 0.863	34.82 / 0.855	35.24 / 0.870

comprehensive view within each contrast shows the fidelity of our model across mask patterns and different amounts of under-sampling.

Figures 5 and 6 provides the qualitative reconstruction results for T1 and FLAIR MRI reconstruction. The T1 target image shows the cerebellum region in the hindbrain while the FLAIR image shows regions around the corpus callosum. The visual results show that (i) KM-MAML is able to recover fine structures much better when compared with other methods for both contrasts and exhibits the least residual error. (ii) MMAML prediction suffers from a blur in the T1 highlighted region as compared to other methods. (iii) The residual images of MAML-based methods show lesser errors than joint training. These observations emphasize the importance of discriminating the representations at the contextual and image levels. The KM hypernetworks capture the semantic relationship between various modes, while the base network is optimized for the image reconstruction task (multi-objective training).

4.2.2. On-the-fly adaptation to unseen multimodal MRI contrasts

To verify that KM-MAML can balance flexibility and robustness, we assess the capabilities of on-the-fly adaptation without fine-tuning to unseen multi-contrast MRI datasets. In this experiment, we consider 24 tasks combining unseen T1, T2, and PD contrasts and unseen acceleration factors, 6x, 7x, 8x, and 9x, with Cartesian and Gaussian mask patterns. Table 3 shows

Figure 4: SSIM plots showing the performance of models for training tasks with varying contrasts, mask types, and acceleration factors (a) T1 (b) T2 (c) PD, and (d) FLAIR respectively. C denotes the Cartesian mask type, and G denotes the Gaussian mask type. For instance, G5 indicates a Gaussian mask pattern with 5x acceleration. Figure (e) shows the consolidated performance of each contrast. FL indicates FLAIR MRI contrast

the performance for twelve of them. Our observations are as follows. (i) Both MMAML and KM-MAML exhibit better generalization than vanilla MAML and joint training. This observation indicates that task-aware modulation is an essential principle in improving the performance of MAML on heterogeneous tasks. (ii) MAML and MMAML show better generalization compared to joint training with respect to SSIM. (iii) KM-MAML can encompass a variety of distribution shifts in contrasts, acceleration factors, and mask patterns with higher improvement margins of over 0.1dB in PSNR and 0.01 SSIM for

Figure 5: Qualitative comparison of the reconstruction performance for T1 MRI. From the left: the ground truth (GT) image with the region of interest (ROI) highlighted as yellow box, GT inset, ZF image, joint training, MAML, MMAML, and KM-MAML. The yellow arrows in the image highlight the improved structure recovery of KMMAML. The residual images with respect to the target indicate that KM-MAML recovers details better than other learning methods.

Figure 6: Qualitative comparison of the reconstruction performance for FLAIR MRI. From the left: the ground truth (GT) image with ROI, GT inset, ZF image, joint training, MAML, MMAML and KM-MAML. As pointed out by yellow arrows, KM-MAML is able to recover structures closer to the target better than other learning methods. This is also evident from the residual images. MMAML and MAML provide better reconstruction quality than SGD but with missing structures around the regions of interest.

most tasks. These observations indicate that the KM hypernetworks exhibit an improved representational capacity to learn discriminative features effectively. At the same time, the hypernetworks learn task-to-task similarities and re-use the information for related tasks with drifts in degradation levels and contrast types.

The box plots in Figure 7 show the comparative study of PSNR and SSIM metrics for the twelve unseen tasks specified in Table 3. We see that joint training, MAML, and MMAML exhibit a drop in performance for PD, T2,

Table 3: Quantitative comparison of SGD, MAML, MMAML, and KM-MAML for on-the-fly adaptation to various unseen tasks combining multiple contrasts, acceleration factors, and mask types with deviated acquisition settings. Green and blue colors indicate the best and the second best metrics respectively. The tasks are denoted in short as contrast type - mask type - acceleration factor

Task	SGD	MAML	MMAML	KM-MAML
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1C6	31.33 / 0.837	31.20 / 0.837	31.23 / 0.832	31.42 / 0.842
T1C7	32.65 / 0.846	32.60 / 0.849	32.50 / 0.843	32.79 / 0.851
T1C8	30.63 / 0.833	30.61 / 0.834	30.60 / 0.830	30.87 / 0.842
T1C9	30.07 / 0.824	30.16 / 0.826	30.08 / 0.822	30.32 / 0.834
T2G6	35.20 / 0.866	34.72 / 0.855	35.10 / 0.869	35.27 / 0.870
T2G7	33.98 / 0.851	33.50 / 0.839	33.82 / 0.855	34.18 / 0.855
T2G8	33.37 / 0.837	33.00 / 0.827	33.23 / 0.838	33.57 / 0.845
T2G9	32.31 / 0.819	31.78 / 0.804	32.06 / 0.820	32.41 / 0.823
PDC6	31.70 / 0.862	31.77 / 0.864	31.70 / 0.867	31.95 / 0.871
PDC7	30.45 / 0.843	30.83 / 0.852	30.86 / 0.855	31.11 / 0.861
PDC8	31.30 / 0.853	31.43 / 0.857	31.53 / 0.862	31.85 / 0.868
PDC9	28.67 / 0.806	28.93 / 0.814	28.81 / 0.812	29.20 / 0.827

and T1, respectively while KM-MAML gives the highest scores in all unseen contrasts. Out of the 24 unseen tasks overall, our method gives the highest scores for 80% of the tasks in terms of PSNR and 92% of the tasks in SSIM.

Figures 8, and 9 show the visual results for the unseen contrasts highlighting the regions around the central lobule of the brain image. In both T1 and PD, the recovery of repeated patterns is much closer with respect to the target image than in other methods. Secondly, in the case of PD (Figure 9), the highlighted region indicated by a yellow arrow in the zero-filled input image (third from left) shows a structure with discontinuity due to aliasing. We notice that other methods still show discontinuity in the image structure while KM-MAML exhibits superior structure recovery. These observations are on par with the least residual error for KM-MAML.

From the quantitative and qualitative results, we see that (i) the meta-initializations of the KM hypernetworks evolve a structure of weights for continuous image generation by extrapolating and interpolating between various contextual settings. (ii) kernel modulation provides a way to re-calibrate the weights by learning global information of multi-modal data and dynamically emphasizing informative features in the base network at each subsampling level. This aspect is very similar to the attention and gating mechanisms [73, 67] applied on CNN features to focus on important regions of the image features.

Figure 7: PSNR and SSIM plots comparing on-the-fly adaptation performance of joint training, MAML, MMAML, and KM-MAML for 12 unseen tasks for T1, T2, and PD MRI data shown in Table 3 with deviations from training data. C denotes the Cartesian mask type, and G denotes the Gaussian mask type.

4.2.3. Adaptation via fine-tuning to unseen multimodal MRI contrasts

The purpose of this experiment (Table 4) is twofold. (i) To compare the adaptation capabilities of the models via fine-tuning to a few gradient steps (ii) To understand whether the weights of the KM hypernetworks are extensible and favorable for rapid adaptation to unseen tasks. We analyze the second aspect with an ablative study of adapting either the base network (columns 4 and 5 in Table 4 and Refer Algorithm 2 in Appendix) or the KM hypernetworks (columns 6 and 7 in Table 4) in MMAML and KM-MAML. A possible scenario wherein adapting the KM hypernetworks would

Figure 8: Qualitative comparison of the on-the-fly adaptation capabilities of the methods to unseen contrasts (T1 MRI). From the left, we have GT image, GT ROI, ZF, joint training, MAML, MMAML, and KM-MAML. The figure shows the superior recovery of a pair of structures in the culmen region of the axial T1 MRI for KM-MAML while other methods exhibit relatively more blur in the region.

Figure 9: Qualitative Results for PD MRI comparing the target image, zero-filled reconstruction, joint training, MAML, MMAML, and KM-MAML with respect to the on-the-fly adaptation capabilities of the models to multiple contrasts. The two region pointed out by the yellow arrows in the predicted images show that KM-MAML exhibits better recovery of fine details over other methods.

be faster and more efficient is when the base network is very deep with more parameters while the hypernetworks are lightweight with relatively lesser weights. We evaluate the adaptation performance on 16 unseen tasks with T1 and PD contrasts with unseen acceleration factors 5x, 6x, 7x, and 9x and unseen mask patterns.

Our observations from the table are as follows: (i) Our first objective is met, wherein KM-MAML exhibits better adaptation over other methods, improving further upon KM (refer to box plots in Figure 10). This observation shows that the proposed model provides stronger mode-specific meta-

Table 4: Quantitative comparison of SGD, MAML, MMAML, and KM-MAML for adaptation via fine-tuning with 10 gradient steps for various unseen contrasts, acceleration factors, and mask types. Ablative studies for adapting either the base or the modulation network for MMAML and KM-MAML

| Task | SGD | MAML | Adapt base n/w | | Adapt modulation n/w | |
| | | | MMAML | KM-MAML | MMAML | KM-MAML |
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1C5	33.77/ .872	33.94/ .876	33.48/ .867	33.67/ .874	33.44/ .866	33.65/ .874
T1C6	31.33/ .837	31.21/ .837	31.30/ .837	31.45/ .844	31.25/ .835	31.43/ .844
T1C7	32.67/ .847	32.62/ .850	32.47/ .843	32.90/ .854	32.39/ .840	32.82/ .851
T1C9	30.08/ .824	30.18/ .827	30.05/ .821	30.34/ .834	30.05/ .821	30.33/ .834
T1G5	35.88/ .852	35.86/ .854	35.75/ .852	36.01/ .858	35.75/ .852	35.97/ .857
T1G6	34.72/ .841	34.57/ .837	34.41/ .834	34.73/ .843	34.26/ .831	34.71/ .841
T1G7	32.45/ .794	32.38/ .792	32.15/ .789	32.56/ .801	32.11/ .788	32.44/ .798
T1G9	30.99/ .779	30.99/ .781	30.83/ .780	30.90/ .785	30.70/ .775	30.89/ .785
PDC5	32.22/ .869	32.54/ .876	32.39/ .875	32.72/ .883	32.37/ .875	32.70/ .883
PDC6	31.71/ .862	31.77/ .864	31.76/ .868	32.00/ .872	31.74/ .867	31.99/ .872
PDC7	30.49/ .844	30.85/ .852	30.90/ .856	31.15/ .861	30.89/ .856	31.15/ .861
PDC9	28.72/ .807	28.95/ .815	28.86/ .813	29.20/ .827	28.83/ .812	29.20/ .827
PDG5	36.46/ .894	36.35/ .895	36.50/ .900	37.10/ .908	36.50/ .900	37.03/ .907
PDG6	33.84/ .852	33.57/ .848	33.87/ .858	34.26/ .865	33.81/ .856	34.25/ .865
PDG7	32.99/ .838	32.72/ .832	32.85/ .840	33.29/ .852	32.72/ .836	33.27/ .851
PDG9	31.50/ .804	31.32/ .804	31.38/ .809	31.91/ .824	31.30/ .805	31.90/ .824

initializations to adapt in a few gradient steps. (ii) Our ablative study shows that KM-MAML outperforms MMAML in both cases of adapting either the base or the modulation network. Also, comparing columns 4 and 7, KM-MAML is better than MMAML. This observation shows that fine-tuning the KM hypernetworks enables adaptation to unseen related tasks in the contextual space while preserving the image reconstruction features learned by the base network. (iii) Adapting the KM hypernetworks gives a competitive performance with adapting the base network.

The qualitative results for adaptation through fine-tuning are shown in Figures 11 and 12. Figures show that KM-MAML provides better reconstructions when compared with other methods. The improvement is consistent as compared to MMAML in both cases of adapting either the base network or the modulation network to unseen tasks.

4.2.4. Interpreting the kernel modulation-based meta-learning

To gain insights into kernel modulation, we analyze KM-MAML by comparing the nearness between the on-the-fly adaptation versus adaptation in few gradient steps. Furthermore, we perform a representational analysis using Centered Kernel Alignment (CKA) [64, 46] metric to understand the abil-

Figure 10: PSNR and SSIM plots comparing generalization through adaptation in few gradient steps for joint training, MAML, MMAML, and KM-MAML for 12 tasks with deviated contrasts, acceleration factors, and mask types for T1, T2, and PD respectively deviated from training data. C denotes the Cartesian mask type, and G denotes the Gaussian mask type. For instance, T2C7 indicates an unseen task with T2 MRI contrast with a Cartesian mask pattern with 7x acceleration.

ity of the kernel modulation network in learning mode-specific discriminative features. The CKA gives a number between 0 and 1 to quantify the correlation between representations. For instance, a CKA value of 0.7 between pre and post modulation features of a base network layer means that 30% of discriminative knowledge associated with a task is induced post-modulation.

Figure 11: Qualitative Results for T1 MRI comparing (from the top left) the target image, ZF image, joint training, MAML, MMAML (pink box), and KM-MAML (green box) with respect to the adaptation capabilities in a few gradient steps for multiple contrasts. The structure pointed out by the yellow arrow shows that the missing structure in the under-sampled input is inadequately captured by other methods while KM-MAML is able to recover the structure with improved accuracy. In the ablative studies, the pink box shows the visual results of MMAML, and the green box shows the results for KM-MAML.

Figure 12: Qualitative Results for PD MRI comparing the target image, ZF image, joint training, MAML, MMAML, and KM-MAML with respect to the adaptation capabilities in a few gradient steps for multiple contrasts. The structure pointed out by the yellow arrow shows that other methods exhibit artifacts present in the ZF image while in KM-MAML the reconstruction is closer to the target. In the ablative studies, the pink box shows the visual results of MMAML, and the green box shows the results for KM-MAML.

4.2.4.1 Comparing on-the-fly adaptation with fine-tuning

We compare the on-the-fly adaptation performance with adaptation via fine-tuning of the KM network to understand the contribution of gradient-based meta-learning for fine-tuning. Table 5 compares on-the-fly adaptation (or the post modulation step) and adaptation via fine-tuning to 10 and 30 gradient steps further after post-modulation. The post-modulation performance improves with adaptation (highlighted in green). The results show that with mode-specific initializations provided by the KM, gradient-based

Table 5: Quantitative comparison of on-the-fly adaptation and adaptation through fine-tuning to 10 and 30 gradient steps. The table shows the adaptation performance for various configurations of T2 and PD MRI contrasts. Note that blue indicates the similarity between on-the-fly and fine-tuning performance. Green showcases the benefits of gradient-based meta-learning by improving the accuracy via fine-tuning in few gradient steps.

TASK	KM-MAML (On-the-fly) PSNR / SSIM	KM-MAML (10 STEPS) PSNR / SSIM	KM-MAML (30 STEPS) PSNR / SSIM
T2C4	36.20 / 0.929	36.25 / 0.928	36.31 / 0.929
T2C5	35.58 / 0.919	35.66 / 0.918	35.69 / 0.919
T2C7	31.73 / 0.857	31.80 / 0.858	31.84 / 0.858
T2C9	30.60 / 0.860	30.61 / 0.859	30.66 / 0.861
T2G4	39.07 / 0.921	39.15 / 0.922	39.19 / 0.923
T2G5	36.66 / 0.889	36.79 / 0.891	36.84 / 0.892
T2G7	34.18 / 0.855	34.27 / 0.857	34.30 / 0.857
T2G8	33.57 / 0.845	33.65 / 0.847	33.70 / 0.848
PDC4	33.95 / 0.904	33.98 / 0.904	33.95 / 0.904
PDC5	32.70 / 0.883	32.71 / 0.883	32.72 / 0.883
PDC6	31.95 / 0.871	32.00 / 0.872	32.02 / 0.873
PDC7	31.11 / 0.861	31.15 / 0.861	31.17 / 0.862
PDG4	39.31 / 0.934	39.34 / 0.935	39.40 / 0.936
PDG5	37.02 / 0.906	36.97 / 0.905	37.16 / 0.910
PDG6	34.23 / 0.864	34.26 / 0.865	34.24 / 0.864
PDG7	33.26 / 0.851	33.29 / 0.852	33.32 / 0.853

meta-learning can further improve the performance on unseen tasks. We also note that the SSIM values post-modulation exhibit closeness to post-adaptation (metrics highlighted in blue), implying that context-aware meta-initializations have gained highly reusable features for new MRI contrasts with deviated acquisition settings.

4.2.4.2 Representational Similarity Analysis

We study the extent to which the base neural network's latent representations (activations) change based on task-specific kernel modulation. Following the recent works [46, 64], we measure the changes in the representations before and after KM based on the CKA metrics to compare similarities in patterns. Figure 13 shows the CKA plots for the three down-sampling layers, the bottleneck layer, and the three upsampling layers of the encoder-decoder base network. The plots show the mean CKA values taken across the 24 unseen MRI tasks. From the plot, our observations are: (i) The CKA metrics are the least, with around 0.2 to 0.3 in the highest levels of the encoder and decoder. This observation shows that these layers, which play a crucial role in recovering fine-grained image details [74] for reconstruction, obtain around

Figure 13: CKA similarity scores between the representations of the seven layers of the encoder-decoder base reconstruction network pre and post kernel modulation. Low CKA similarity at the top layers of the encoder and decoder shows that these layers must change significantly to recover structures according to different modes of the MRI contrasts. The CKA similarity gradually increases with its peak in the latent layer which shows that low frequency details of related multi-contrast images exhibit hight similarity. In the figure *conv_down_0* and *conv_up_2* are the top most levels of the encoder-decoder network. *conv_down_1* and *conv_up_1* form the second sub-sampling level and so on. The *latent_layer* is the bottleneck convolution layer. Note that the variations in the box plots are small indicating that hill-like profile is consistent across all the unseen contrasts (T1, T2, and PD).

70 to 80% mode-specific discriminative knowledge post KM. (ii) Representational similarity increases over sub-sampling levels and reaches a maximum of about 0.78 at the bottleneck layer from the encoder. This observation indicates that lower resolution layers capture low-frequency details that are mode-invariant. Each KM hypernetwork of the corresponding base layer takes varying roles of learning mode-specific and mode-invariant knowledge wherein the top layers learn more mode-specific features while the bottom layer learns more mode-invariant features. (iii) CKA values gradually decrease in the decoder layers as we move towards the output layer to around 28%. This observation is very similar to the CKA analysis in Almost no inner loop (ANIL)[46] for classification tasks wherein the head of the network exhibits the least CKA similarity while the intermediate layers exhibit higher CKA scores. This analysis provides an architecture-level interpretation of kernel modulation of the base network.

Figure 14: Qualitative results for the ablative study. Top: From the left are ZF image, and the predictions of U-Net, U-Net + GBML, and U-Net + KM + GBML. Bottom: From the left, target inset, residual error images with respect to the target for ZF, U-Net, U-Net + GBML, and U-Net + KM + GBML

Table 6: Ablative study showing the role of the base network alone with SGD (single level of optimization), base network alone with gradient-based meta-learning which has two levels of optimization, and the proposed model with base network and kernel modulation network as the architecture and gradient-based meta-learning as the learning process. T1 denotes T1 MRI contrast, C - Cartesian mask pattern, and G - Gaussian mask pattern. For example, T1C8 indicates T1 MRI image with Cartesian mask under-sampling at 5x acceleration.

Task	U-Net w/o GBML	U-Net + GBML	U-Net + KM + GBML
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1C4	36.09 / 0.889	36.45 / 0.900	36.75 / 0.915
T1C5	34.78 / 0.861	35.22 / 0.875	35.58 / 0.895
T1C8	32.52 / 0.817	32.99 / 0.839	33.37 / 0.864
T1G4	40.20 / 0.913	40.59 / 0.922	41.02 / 0.935
T1G5	38.20 / 0.874	38.67 / 0.887	39.05 / 0.906
T1G8	34.85 / 0.802	35.33 / 0.820	35.96 / 0.859

4.2.5. Ablative study of the model and the learning process

We perform an ablative study to understand the role of meta-learning and kernel modulation. We consider 6 acquisition contexts with T1 MRI images, Cartesian and Gaussian mask patterns, and acceleration factors 4x, 5x, and 8x. Table 6 and Figure 14 show the quantitative and qualitative results, respectively, for the ablative study considering three cases, 1) Only the base network with conventional joint training involving a single level of optimization, 2) Only the base network trained with gradient-based meta-learning

Table 7: Quantitative comparison with context-specific deep cascaded MRI reconstruction architectures for fixed anatomy under study (cardiac), acceleration factors 4x and 5x, and Cartesian mask type. Green indicates that our model can operate on multiple acquisition contexts while matching the performance of context-specific models.

Method	4x PSNR / SSIM	5x PSNR / SSIM
ZF	24.27 / 0.699	23.82 / 0.674
DAGAN [55]	28.52 / 0.841	28.02 / 0.825
DC-CNN [56]	32.75 / 0.920	31.75 / 0.905
DC-DEN [65]	33.22 / 0.925	32.30 / 0.913
DC-RDN [66]	32.95 / 0.923	32.09 / 0.912
DC-UNet [58]	33.17 / 0.928	32.55 / 0.919
MICCAN [67]	33.38 / 0.930	32.52 / 0.918
MAC-ReconNet [8]	32.98 / 0.923	32.07 / 0.910
OUCR [68]	32.98 / 0.923	32.07 / 0.910
ViT [69]	28.55 / 0.828	27.86 / 0.806
SWIN [70]	30.15 / 0.869	29.24 / 0.848
SFT-KD-Recon [8]	32.03 / 0.907	30.93 / 0.888
KM-MAML	33.40 / 0.930	32.64 / 0.919

involving two-level optimization at the task level, and 3) With both base and kernel modulation networks within the gradient-based meta-learning process. From the results, it is clear that under variations in the contexts, GBML can provide better meta-initializations as compared to conventional joint training. Furthermore, combining model-based meta-learning via the context-specific kernel modulation using hypernetworks and optimization-based meta-learning enhances the learning with improvement margins of $\sim$1 dB in PSNR and $\sim$0.01 in SSIM.

5. Model-based Comparative Studies

5.0.1. Comparison against context-specific MRI reconstruction architectures

We provide architecture-based performance comparisons to show that our method can operate in multiple acquisition context-based configurations as well as match the reconstruction performance of other context-specific networks. We perform kernel modulation using a single set of KM hypernetworks on the base network (U-Net) in deep cascaded mode [56]. We compare our network against DAGAN, and various deep cascaded CNNs that are trained for a specific context - cardiac anatomy, a fixed Cartesian under-sampling pattern, and a specific acceleration factor (4x or 5x). We train the adaptive MRI reconstruction models, MAC-ReconNet and KM-MAML on ten acquisition configurations with the cardiac anatomy, varying mask types - Gaussian and Cartesian, and varying acceleration factors - 2x, 3.3x, 4x, 5x, and 8x.

Figure 15: Qualitative comparison of KM-MAML in deep-cascaded mode against ZF, context-specific and adaptive MRI reconstruction architectures. The yellow arrows in the figures point below the ventricle region of the cardiac MRI anatomy with a heavy aliasing on the ZF image. KM-MAML recovers finer details much closer to the target as compared to other methods where the aliasing artifacts are still present.

Table 8: Comparative study of scalability of context-specific networks (DC-Unet and DC-DEN) and the proposed network when combining multiple anatomies with different image resolution levels, mask patterns, and acceleration factors. T1 denotes T1 contrast, and PD denotes proton density-weighted contrast. B and K denote brain and knee anatomies. C and G denote Cartesian and Gaussian mask patterns. For example, the task PDKC5 denotes the context with PD knee, Cartesian, 5x acceleration.

Task	DC-Unet	DC-DEN	KM-MAML
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1BC5	42.67 / 0.984	42.87 / 0.987	43.64 / 0.990
T1BC8	38.81 / 0.967	38.25 / 0.967	39.33 / 0.974
T1BG5	49.10 / 0.990	50.48 / 0.996	51.50 / 0.998
T1BG8	46.27 / 0.986	46.44 / 0.991	47.99 / 0.994
PDKC5	35.64 / 0.921	35.39 / 0.916	35.83 / 0.923
PDKC8	32.81 / 0.877	32.18 / 0.865	32.96 / 0.880
PDKG5	37.42 / 0.935	37.20 / 0.929	37.82 / 0.940
PDKG8	34.58 / 0.895	33.00 / 0.869	35.00 / 0.902

Table 7 shows that our model can operate in multiple configurations and outperform the context-specific models in terms of PSNR and SSIM. We also note that as compared to MAC-ReconNet our model performs better when trained in multiple contexts. Unlike our network which has both context-specific weights provided by the KM hypernetworks and context-invariant weights of the base network, MAC-ReconNet lacks context-invariant weights of the base network, as all the weights of the base network are predicted by the hypernetworks. The qualitative results (Figure. 15) reveal that KM-MAML is able to reconstruct images closer to the target image as compared to other methods.

5.0.2. Multiple Acquisition Contexts - Multiple Anatomies and Image Resolution levels

As the main goal of this work is to enable more capabilities to approximate diverse contexts, we compare the scalability of our network with other networks when combining multiple anatomies, contrasts, and image resolution levels (320×320 and 240×240), mask pattern and acceleration factors, making to 12 contexts in a single training. For our comparative study, we take two of the top-performing architectures from Table 7, the DC-Unet (encoder-decoder CNN like MICCAN) and DC-DEN (DenseNet CNN configuration). Table 8 shows the comparative study of the scalability of these models and our method to multiple acquisition contexts. Under diverse data settings, the modulated weights in the proposed method provide context-adaptive reconstruction and exhibit better metrics than DC-UNet and DC-DEN.

Figure 16: Qualitative comparison of KM-MAML with DC-DEN and DC-UNet when combining multiple anatomies, mask patterns, and acceleration factors. The yellow arrows pointing at the medial meniscus and ligament regions of the knee MRI anatomy show that KM-MAML reconstructs the image much closer to the target as compared to DC-DEN and DC-UNet under heterogeneous data scenario.

The improvement margins over DC-UNet are $\sim$ 0.8 dB in PSNR and $\sim$ 0.006 in SSIM, while the improvement margins over DC-DEN are $\sim$ 1 dB in PSNR and $\sim$ 0.01 in SSIM. We believe the main reason for this observation is that DC-DEN and DC-UNet learn a single set of shared weights that pull the reconstruction towards an average of the possible reconstructions equidistant from all diverse acquisition contexts considered at train time. These weights are inadequate to scale to multiple diverse contexts. On the other hand, KM-MAML exhibits dynamic context-specific weights at inference time. The qualitative results for the knee context in Figure 16 show that the reconstructed image corresponding to DC-DEN shows missing structures and that corresponding to DC-UNet shows smudged regions, while the predicted image of KM-MAML is much closer to the target image.

5.0.3. Fine-tuning Vs. Transfer learning with Context-specific networks

In this experiment, we have taken the models trained on multiple anatomies and contrasts (proton density knee 320×320 and T1 brain 240×240), mask patterns, and acceleration factors (as shown in Section 5.0.2) as the pre-trained models. The models are fine-tuned (one or two epochs i.e. visiting the samples once or a maximum of two times) and transfer learned (ten epochs) to unseen contexts covering cardiac anatomy with Cartesian, radial (unseen mask pattern) and Gaussian mask patterns with acceleration factors 3.3x, 5x, and 10x and image resolution level of 150×150. From the quantitative results shown in Table 9, our observations are as follows. 1) The performance of KM-MAML after adaptation is, in general, better than the

Table 9: Comparative study of context-specific networks (DC-Unet and DC-DEN) and the proposed network when transfer learning from knee 320×320 and brain 240×240 contexts to cardiac 150×150 with Gaussian, radial and Cartesian mask patterns and acceleration factors 3.3x, 5x and 10x. CC, CR, and CG indicate the cardiac anatomy with Cartesian, radial, and Gaussian mask under-sampling patterns, respectively. For instance, task CR5 denotes cardiac anatomy undersampled with radial masks with 5x under-sampling. The numbers highlighted in green show that the fine-tuning performance (Epoch 1 and 2) of KM-MAML matches the transfer learned performance (Epoch 10) of context-specific networks.

Task	DC-DEN		DC-UNet		KM-MAML		
	Epoch 1 PSNR/ SSIM	Epoch 10 PSNR/ SSIM	Epoch 1 PSNR/ SSIM	Epoch 10 PSNR/ SSIM	Epoch 1 PSNR/ SSIM	Epoch 2 PSNR/ SSIM	Epoch 10 PSNR/ SSIM
CC3.3	33.58/ .934	34.30/ .941	34.24/ .943	34.61/ .946	34.43/ .945	34.58/ .946	34.82/ .948
CC5	30.73/ .888	31.60/ .903	31.83/ .909	32.23/ .915	32.01/ .910	32.16/ .913	32.45/ .917
CC10	24.98/ .727	25.54/ .749	25.93/ .765	26.31/ .778	26.12/ .770	26.26/ .775	26.49/ .783
CR3.3	37.47/ .961	38.31/ .966	38.22/ .967	38.90/ .971	38.71/ .970	38.90/ .971	39.18/ .973
CR5	33.86/ .923	34.66/ .933	34.72/ .936	35.14/ .941	34.92/ .939	35.07/ .941	35.34/ .944
CR10	27.95/ .794	28.64/ .816	29.02/ .832	29.43/ .842	29.15/ .834	29.33/ .839	29.57/ .846
CG3.3	39.16/ .975	39.94/ .978	39.85/ .978	40.34/ .980	40.29/ .980	40.42/ .981	40.62/ .982
CG5	35.89/ .951	36.63/ .957	36.72/ .958	37.27/ .962	37.05/ .961	37.21/ .962	37.46/ .963
CG10	29.20/ .845	30.21/ .868	30.73/ .880	31.19/ .888	30.89/ .880	31.03/ .884	31.37/ .890

other two methods in both fine-tuning (Epoch 1 and 2) and transfer learning cases (Epoch 10) for all mask patterns and acceleration factors. 2) When the fine-tuning (Epoch 1 and Epoch 2) performance of KM-MAML is compared with DC-UNet, and DC-DEN after transfer learning (Epoch 10), we note that KM-MAML is competitive with DC-UNet models for 5 out of 8 cases at epoch 1 and almost all the cases at epoch 2. These cases are highlighted in green in Table 9. 3) KM-MAML fine-tuning shows significant improvement over DC-DEN after transfer learning. These observations reveal superior meta-learning capabilities of KM-MAML with context-specific learning using kernel modulation as compared to other models. The qualitative results in Figure 17 comparing the transfer learning performance of DC-DEN and DC-UNet with the Epoch 2 fine-tuning performance of KM-MAML show that our model has the capabilities to generalize via faster adaptation to

GT Image   DC-DEN   DC-UNet   KM-MAML   GT inset
PSNR / SSIM   34.09 / 0.959   34.51 / 0.963   35.14 / 0.968

Figure 17: Qualitative comparison of KM-MAML with DC-DEN and DC-UNet when adapting from knee and brain to cardiac contexts with 5x acceleration. The yellow arrows pointing at the ventricle regions show that KM-MAML is able to recover fine structures and textures with less blur than other methods.

unseen data domains. The visual quality of KM-MAML prediction is much closer to the target with respect to the recovered textures and patterns as compared to DC-DEN and DC-UNet.

5.0.4. Comparative Study using large scale clinical data

We demonstrate our comparative study of the transfer learning performance of DC-UNet and the on-the-fly adapted performance of KM-MAML on the fastMRI dataset [21], a large-scale collection of clinical MR images. The dataset consists of coronal PD and PDFS images with varying Cartesian mask patterns for 4x and 8x accelerations. The validation set consists of mask patterns and varying image resolution levels different from the training dataset. The quantitative results for DC-UNet are PSNR / SSIM: 31.11 / 0.852, and DC-UNet + transfer learned to 10 epochs are, PSNR / SSIM: 31.14 / 0.854. The quantitative results for KM-MAML in the one-the-fly adaptation setting (i.e. without fine-tuning) are, PSNR / SSIM: 31.16 / 0.855. The qualitative results are shown in Figure 18. The results demonstrate the consistency in generalization capabilities of dynamic weight prediction and kernel modulation of the base network across varying mask patterns and resolution levels when trained on large-scale training data.

6. Summary and Conclusion

In this work, we adaptively learn the structure of multimodal data distributions using KM-MAML, a learning model that combines the strengths of model-based evolutionary deep learning and optimization-based meta-learning. The hypernetworks output low-rank approximation weights to

Figure 18: Qualitative comparison of KM-MAML with DC-UNet when trained on large scale data (fastMRI) and adapting to few samples of images under-sampled on unseen mask patterns and image resolution levels of knee MRI.

modulate the base network based on different modalities of the acquired image data. The training objectives include learning the inductive bias of multiple modes of heterogeneous data and low-level image features. To achieve the multi-objective training, our model optimally utilizes the hypernetworks via kernel modulation and gradient-based meta-learning (Figure 2 and 3).

We have demonstrated the efficacy of our method for MC-MRI reconstruction considering the two essential fastMRI research directions: multimodal and transfer learning across MRI scanners. We considered benchmark MC-MRI datasets, such as MRBrainS, SRI24 Atlas, and IXI datasets, to compare the scalability of the learning methods and the cardiac ACDC dataset, and fastMRI knee datasets to compare the reconstruction performance against various acquisition context-specific MRI reconstruction networks. Scalability to numerous multimodal acquisition settings for clinical use is facilitated through hypernetworks (Figure 4) that impart the necessary context-aware bias into base CNN and is further enhanced by discriminatively fine-tuning to the shifted target distribution. Representation similarity analysis pro-

vides insights into mode-specific knowledge in the high-resolution layers of the base network (Figure 13). Our proposed KM-based architecture outperforms various context-specific MRI reconstruction architectures (Tables 7 and 8) and the adaptive MRI reconstruction architecture, MAC-ReconNet in the multiple context-based setting quantitatively and qualitatively.

Our future research directions include analyzing the model in terms of better context embedding, extending the work for multi-coil reconstruction with self-supervision, other regression tasks like image imputation in MRI, and contrast augmentation techniques to improve the generalization capabilities of the model.

7. Acknowledgement

References

[1] J. M. J. Valanarasu, P. Guo, V. VS, V. M. Patel, On-the-fly test-time adaptation for medical image segmentation (2022). `doi:10.48550/ ARXIV.2203.05574`.

[2] R. C. Hien Van Nguyen, Ronald Summers, Metalearning with Medical Imaging and Health Informatics Applications, The Elsevier and MICCAI Society book Series, 2022.

[3] Y. Ma, S. Zhao, W. Wang, Y. Li, I. King, Multimodality in meta-learning: A comprehensive survey, Knowledge-Based Systems 250 (2022) 108976. `doi:https://doi.org/10.1016/j.knosys.2022. 108976`.

[4] J. Requeima, J. Gordon, J. Bronskill, S. Nowozin, R. E. Turner, Fast and Flexible Multi-Task Classification Using Conditional Neural Adaptive Processes, Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS), 2019.

[5] D. Chen, Q. Fan, J. Liao, A. Aviles-Rivero, L. Yuan, N. Yu, G. Hua, Controllable Image Processing via Adaptive FilterBank Pyramid, IEEE Trans. Image Process. 29 (2020) 8043–8054. `doi:10.1109/TIP.2020. 3009844`.

[6] Q. Fan, D. Chen, L. Yuan, G. Hua, N. Yu, B. Chen, A General Decoupled Learning Framework for Parameterized Image Operators, IEEE Trans. Pattern Anal. Mach. Intell. (2019) 1–1`doi:10.1109/TPAMI. 2019.2925793`.

[7] X. Liu, J. Wang, F. Liu, S. K. Zhou, Universal undersampled mri reconstruction, in: M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, Springer International Publishing, Cham, 2021, pp. 211–221.

[8] S. Ramanarayanan, B. Murugesan, K. Ram, M. Sivaprakasam, MAC-ReconNet: A Multiple Acquisition Context based Convolutional Neural Network for MR Image Reconstruction using Dynamic Weight Prediction, in: MIDL, Vol. 121, 2020, pp. 696–708.

[9] R. Vilalta, Y. Drissi, A perspective view and survey of meta-learning, Artificial Intelligence Review 18 (2) (2002) 77–95. `doi:10.1023/A: 1019956318069`.

[10] C. Finn, P. Abbeel, S. Levine, Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, in: Proc. ICML, Vol. 70, 2017, pp. 1126–1135.

[11] F. Zhou, B. Wu, Z. Li, Deep meta-learning: Learning to learn in the concept space, CoRR abs/1802.03596 (2018). `arXiv:1802.03596`.

[12] R. Subramanyam, M. Heimann, J. Thathachar, R. Anirudh, J. J. Thiagarajan, Contrastive knowledge-augmented meta-learning for few-shot classification (2022). `doi:10.48550/ARXIV.2207.12346`.

[13] R. Vuorio, S.-H. Sun, H. Hu, J. J. Lim, Multimodal model-agnostic meta-learning via task-aware modulation, in: Neural Information Processing Systems, 2019.

[14] Y. Bi, B. Xue, P. Mesejo, S. Cagnoni, M. Zhang, A survey on evolutionary computation for computer vision and image analysis: Past, present, and future trends, IEEE Transactions on Evolutionary Computation (2022) 1–1`doi:10.1109/TEVC.2022.3220747`.

[15] Y. Bi, B. Xue, P. Mesejo, S. Cagnoni, M. Zhang, A survey on evolutionary computation for computer vision and image analysis: Past, present, and future trends, IEEE Transactions on Evolutionary Computation (2022) 1–1`doi:10.1109/tevc.2022.3220747`.

[16] T. M. Hospedales, A. Antoniou, P. Micaelli, A. J. Storkey, Meta-learning in neural networks: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence (2021) 1–1`doi:10.1109/TPAMI.2021.3079209`.

[17] D. Ha, A. Dai, Q. V. Le, Hypernetworks, arXiv preprint arXiv:1609.09106 (2016).

[18] S. Yantis, The neural basis of selective attention, Current Directions in Psychological Science 17 (2) (2008) 86–90. `doi:10.1111/j.1467-8721.2008.00554.x`.

[19] M. Przewięźlikowski, P. Przybysz, J. Tabor, M. Zięba, P. Spurek, Hyper-maml: Few-shot adaptation of deep models with hypernetworks (2022). `doi:10.48550/ARXIV.2205.15745`.

[20] J. Wang, Q. Yang, Q. Yang, L. Xu, C. Cai, S. Cai, Joint optimization of cartesian sampling patterns and reconstruction for single-contrast and multi-contrast fast magnetic resonance imaging, Computer Methods and Programs in Biomedicine 226 (2022) 107150. `doi:https://doi.org/10.1016/j.cmpb.2022.107150`.

[21] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, Y. W. Lui, fastMRI: An open dataset and benchmarks for accelerated MRI, ArXiv abs/1811.08839 (2018).

[22] L. Sun, Z. Fan, X. Fu, Y. Huang, X. Ding, J. Paisley, A Deep Information Sharing Network for Multi-Contrast Compressed Sensing MRI Reconstruction, IEEE Transactions on Image Processing 28 (12) (2019) 6141–6153. `doi:10.1109/TIP.2019.2925288`.

[23] B. Zhou, S. K. Zhou, Dudornet: Learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior, in: 2020 IEEE/CVF

Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 4272–4281. `doi:10.1109/CVPR42600.2020.00433`.

[24] B. Zhou, N. Dey, J. Schlemper, S. S. M. Salehi, C. Liu, J. S. Duncan, M. Sofka, Dsformer: A dual-domain self-supervised transformer for accelerated multi-contrast mri reconstruction (2022). `doi:10.48550/ARXIV.2201.10776`.

[25] Y. Tian, X. Zhao, W. Huang, Meta-learning approaches for learning-to-learn in deep learning: A survey, Neurocomputing 494 (2022) 203–223. `doi:https://doi.org/10.1016/j.neucom.2022.04.078`.

[26] A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, Q. Zhang, Multiobjective evolutionary algorithms: A survey of the state of the art, Swarm and Evolutionary Computation 1 (1) (2011) 32–49. `doi:https://doi.org/10.1016/j.swevo.2011.03.001`.

[27] X. Song, W. Gao, Y. Yang, K. Choromanski, A. Pacchiano, Y. Tang, Esmaml: Simple hessian-free meta learning, in: International Conference on Learning Representations, 2020.

[28] M. Jeon, Y. Noh, K. Jeon, S. Lee, I. Lee, Data gap analysis of ship and maritime data using meta learning, Applied Soft Computing 101 (2021) 107048. `doi:https://doi.org/10.1016/j.asoc.2020.107048`.

[29] A. Fukui, D. H. Park, D. Yang, A. Rohrbach, T. Darrell, M. Rohrbach, Multimodal compact bilinear pooling for visual question answering and visual grounding, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 457–468. `doi:10.18653/v1/D16-1044`.

[30] J. Yu, J. Li, Z. Yu, Q. Huang, Multimodal transformer with multi-view visual representation for image captioning, IEEE Transactions on Circuits and Systems for Video Technology 30 (12) (2020) 4467–4480. `doi:10.1109/TCSVT.2019.2947482`.

[31] C. Fernando, D. Banarse, M. Reynolds, F. Besse, D. Pfau, M. Jaderberg, M. Lanctot, D. Wierstra, Convolution by evolution: Differentiable pattern producing networks, in: Proceedings of the Genetic and

Evolutionary Computation Conference 2016, GECCO '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 109–116. `doi:10.1145/2908812.2908890`.

[32] J. von Oswald, C. Henning, J. Sacramento, B. F. Grewe, Continual learning with hypernetworks, CoRR abs/1906.00695 (2019). `arXiv:1906.00695`.

[33] K. Frans, O. Witkowski, Population-based evolution optimizes a meta-learning objective, CoRR abs/2103.06435 (2021). `arXiv:2103.06435`.

[34] J. Lehman, K. O. Stanley, Improving evolvability through novelty search and self-adaptation, in: 2011 IEEE Congress of Evolutionary Computation (CEC), 2011, pp. 2693–2700. `doi:10.1109/CEC.2011.5949955`.

[35] K. O. Stanley, D. B. D'Ambrosio, J. Gauci, A hypercube-based encoding for evolving large-scale neural networks, Artificial Life 15 (2) (2009) 185–212. `doi:10.1162/artl.2009.15.2.15202`.

[36] B. De Brabandere, X. Jia, T. Tuytelaars, L. Van Gool, Dynamic Filter Networks, in: NeurIPS, 2016, p. 667–675.

[37] S. Ye, D. Chen, S. Han, Z. Wan, J. Liao, Meta-PU: An Arbitrary-Scale Upsampling Network for Point Cloud, IEEE Transactions on Visualization and Computer Graphics (2021) 1–1`doi:10.1109/TVCG.2021.3058311`.

[38] C. Finn, S. Levine, Meta-learning and universality: Deep representations and gradient descent can approximate any learning algorithm, CoRR abs/1710.11622 (2017). `arXiv:1710.11622`.

[39] W. Wang, R. Guo, Y. Tian, W. Yang, CFSNet: Toward a Controllable Feature Space for Image Restoration, in: IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 4139–4148. `doi:10.1109/ICCV.2019.00424`.

[40] S. Aharon, G. Ben-Artzi, Hyperres: Efficient hypernetwork-based continuous image restoration (2022). `doi:10.48550/ARXIV.2206.05970`.

[41] C. Finn, P. Abbeel, S. Levine, Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks, in: Proc. ICML, Vol. 70, 2017, pp. 1126–1135.

[42] H. S. Behl, A. G. Baydin, P. H. S. Torr, Alpha MAML: adaptive model-agnostic meta-learning, CoRR abs/1905.07435 (2019). `arXiv:1905.07435`.

[43] S. Baik, S. Hong, K. M. Lee, Learning to forget for meta-learning, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2376–2384. `doi:10.1109/CVPR42600.2020.00245`.

[44] M. A. Jamal, G.-J. Qi, Task agnostic meta-learning for few-shot learning, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11711–11719. `doi:10.1109/CVPR.2019.01199`.

[45] J. Zhu, J. Jang-Jaccard, I. Welch, H. Al-Sahaf, S. Camtepe, A ransomware triage approach using a task memory based on meta-transfer learning framework (2022). `doi:10.48550/ARXIV.2207.10242`.

[46] A. Raghu, M. Raghu, S. Bengio, O. Vinyals, Rapid learning or feature reuse? towards understanding the effectiveness of MAML, CoRR abs/1909.09157 (2019). `arXiv:1909.09157`.

[47] S. Baik, J. Choi, H. Kim, D. Cho, J. Min, K. M. Lee, Meta-learning with task-adaptive loss function for few-shot learning, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 9445–9454. `doi:10.1109/ICCV48922.2021.00933`.

[48] S. Baik, M. Choi, J. Choi, H. Kim, K. M. Lee, Meta-learning with adaptive hyperparameters, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, Vol. 33, Curran Associates, Inc., 2020, pp. 20755–20765.

[49] Z. Wang, H. Zhang, Z. Cheng, B. Chen, X. Yuan, Metasci: Scalable and adaptive reconstruction for video compressive sensing, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 2083–2092. `doi:10.1109/CVPR46437.2021.00212`.

[50] L. Zintgraf, K. Shiarli, V. Kurin, K. Hofmann, S. Whiteson, Fast context adaptation via meta-learning, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine

Learning, Vol. 97 of Proceedings of Machine Learning Research, PMLR, 2019, pp. 7693–7702.

[51] I. Skorokhodov, S. Ignatyev, M. Elhoseiny, Adversarial generation of continuous images, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 10748–10759. `doi:10.1109/CVPR46437.2021.01061`.

[52] A. Q. Wang, A. V. Dalca, M. Sabuncu, Regularization-Agnostic Compressed Sensing MRI Reconstruction with Hypernetworks, ArXiv abs/2101.02194 (2021).

[53] A. Q. Wang, A. V. Dalca, M. R. Sabuncu, Computing multiple image reconstructions with a single hypernetwork (2022). `doi:10.48550/ARXIV.2202.11009`.

[54] S. M. Jayakumar, J. Menick, W. M. Czarnecki, J. Schwarz, J. W. Rae, S. Osindero, Y. W. Teh, T. Harley, R. Pascanu, Multiplicative interactions and where to find them, in: International Conference on Learning Representations, 2020.

[55] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, D. Firmin, DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction, IEEE Trans. Med. Imaging 37 (6) (2018) 1310–1321. `doi:10.1109/TMI.2017.2785879`.

[56] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, D. Rueckert, A Deep Cascade of Convolutional Neural Networks for MR Image Reconstruction, in: Proc. Inf. Process. Med. Imaging, 2017, pp. 647–658.

[57] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2015, pp. 234–241.

[58] L. Sun, Z. Fan, X. Ding, Y. Huang, J. Paisley, Joint CS-MRI Reconstruction and Segmentation with a Unified Deep Network, in: Proc. Inf. Process. Med. Imaging, 2019, pp. 492–504.

[59] A. M. Mendrik, K. L. Vincken, H. J. Kuijf, M. Breeuwer, W. H. Bouvy, J. De Bresser, A. Alansary, M. De Bruijne, A. Carass, A. El-Baz, et al.,

MRBrainS challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI scans, Comput. Intell. Neurosci. 2015 (2015) 1.

[60] T. Rohlfing, N. Zahr, E. Sullivan, A. Pfefferbaum, The SRI24 multichannel atlas of normal adult human brain structure, Human brain mapping 31 (5) (2010) 798–819. `doi:10.1002/hbm.20906`.

[61] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, P. Jodoin, Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?, IEEE Trans. Med. Imaging 37 (11) (2018) 2514–2525.

[62] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, F. Knoll, Learning a Variational Network for Reconstruction of Accelerated MRI Data, Magn. Reson. Med. 79 (6) (2018) 3055–3071. `arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.26977`, `doi:10.1002/mrm.26977`.

[63] S. Ramanarayanan, B. Murugesan, K. Ram, M. Sivaprakasam, DC-WCNN: A Deep Cascade of Wavelet Based Convolutional Neural Networks for MR Image Reconstruction, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 1069–1073. `doi:10.1109/ISBI45749.2020.9098491`.

[64] S. Kornblith, M. Norouzi, H. Lee, G. Hinton, Similarity of neural network representations revisited, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, Vol. 97 of Proceedings of Machine Learning Research, PMLR, 2019, pp. 3519–3529.

[65] H. Wu, Y. Wu, L. Sun, C. Cai, Y. Huang, X. Ding, A deep ensemble network for compressed sensing mri, in: Proc. Neural Inf. Process., 2018, pp. 162–171.

[66] L. Sun, Z. Fan, Y. Huang, X. Ding, J. Paisley, Compressed sensing mri using a recursive dilated network, in: Proc. AAAI Conf. Artif. Intell., 2018.

[67] Q. Huang, D. Yang, P. Wu, H. Qu, J. Yi, D. Metaxas, MRI Reconstruction Via Cascaded Channel-Wise Attention Network, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 1622–1626. doi:10.1109/ISBI.2019.8759423.

[68] P. Guo, J. M. J. Valanarasu, P. Wang, J. Zhou, S. Jiang, V. M. Patel, Over-and-under complete convolutional rnn for mri reconstruction, Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention 12906 (2021) 13–23.

[69] K. Lin, R. Heckel, Vision transformers enable fast and robust accelerated MRI, in: Medical Imaging with Deep Learning, 2022.

[70] J. Huang, Y. Fang, Y. Wu, H. Wu, Z. Gao, Y. Li, J. Del Ser, J. Xia, G. Yang, Swin transformer for fast mri, Neurocomputing 493 (2022) 281–304.

[71] M. N. Gayathri, S. Ramanarayanan, M. A. Fahim, R. G. S, K. Ram, M. Sivaprakasam, Sft-kd-recon: Learning a student-friendly teacher for knowledge distillation in magnetic resonance image reconstruction (2023). arXiv:2304.05057.

[72] B. Murugesan, S. Ramanarayanan, S. Vijayarangan, K. Ram, N. R. Jagannathan, M. Sivaprakasam, A deep cascade of ensemble of dual domain networks with gradient-based t1 assistance and perceptual refinement for fast mri reconstruction, Computerized Medical Imaging and Graphics 91 (2021) 101942. doi:https://doi.org/10.1016/j.compmedimag.2021.101942.

[73] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141. doi:10.1109/CVPR.2018.00745.

[74] J. C. Ye, W. K. Sung, Understanding Geometry of Encoder-Decoder CNNs, in: Proc. ICML, Vol. 97, 2019, pp. 7064–7073.

Appendix A. Appendix

Appendix A.1. Data Fidelity Operation in MRI reconstruction

The proposed architecture of KM-MAML has a data fidelity (DF) block [56] in k-space domain after the CNN base network to ensure that the CNN reconstruction is consistent with the acquired k-space measurements. The data fidelity operation f_{df} can be expressed as,

$$
\hat{x}_{df} = \begin{cases} \hat{x}_{CNN}(k) & k \notin \Omega \\ \frac{\hat{x}_{CNN}(k) + \lambda \hat{x}_u(k)}{1+\lambda} & k \in \Omega \end{cases} \tag{A.1}
$$

Here, $\hat{x}_{CNN} = F_f x_{CNN}$, $\hat{x}_u = F_f x_u$, Ω is the index set of sampled k-space data, F_f is the Fourier encoding matrix, and $\hat{x}_{df}$ is the corrected k-space and the data fidelity weight $\lambda \rightarrow \infty$. The reconstructed image is obtained by inverse Fourier encoding of $\hat{x}_{df}$, i.e. $x_{df} = F_f^H \hat{x}_{df}$ (Figure A.19).

Appendix A.2. Glossary of Technical Terms

Meta-initialization or meta-parameters θ - globally shared initialization point of parameters (weights) of a task-oriented base neural network trained on various tasks (for example, different classes in classification tasks) using gradient-based meta-learning, such that a few gradient steps from the initialization parameters can generalize to new related tasks (new classes in classification tasks).

Task-specific parameters ϕ_M - In meta-learning, task-specific parameters refer to the parameters of a neural network that are obtained after rapidly adapting the network to a specific task M encountered during the inner loop of the meta-testing phase. These parameters are different from the meta-initialization parameters that are optimized in the outer loop across tasks [10].

Multimodal MRI data - Multimodal data refers to the data collected through different acquisition technologies. The output of each acquisition technology is represented as a mode in the form of a dataset associated with a medium of expression, such as vision, audio or text. Multimodal MRI data refers to the use of multiple imaging techniques within a single MRI examination to gather complementary information about the structure and function of interest.

Hypernetworks - Hypernetworks are used to adaptively generate weights to initialize or update the parameters of another network, called a base or

Data fidelity (DF) Operation f_{df} for k-space consistency

F_f - Fourier encoding matrix

F_f^H - Inverse Fourier encoding matrix

Figure A.19: K-space Data fidelity for MRI reconstruction

backbone network [17]. Hypernetworks form the basis for model-based meta-learning [16].

Inductive bias - Inductive bias refers to the prior knowledge or assumptions that a learning algorithm uses to make predictions or generalize from training data to unseen data. The meta-learning process learns the inductive bias in the form of the meta-initializations, which allow the model to generalize by adapting quickly to new tasks with limited data [50].

Mode-specific inductive bias - Meta-initializations closer to the given target data mode in multimodal data.

Rank-1 kernel modulation - Element-wise multiplication of the base network kernels with the predicted weights of the hypernetwork. The predicted weights are two vectors corresponding to the number of input channels C_{in} and output channels C_{out} of the corresponding base network layer. For example, if the base network layer has $C_{out} \times C_{in}$ kernels each of size $k \times k$, then the hypernetwork predicts two vectors of sizes $C_{out} \times 1$ and $1 \times C_{in}$, such that the outer product of these two vectors gives a matrix of size $C_{out} \times C_{in}$. The hypernetwork predicts $C_{out} + C_{in}$ weights rather than $C_{out} \times C_{in}$ weights to modulate each kernel of the convolution layer differently.

Table A.10: Quantitative comparison of SGD, MAML, MMAML and KM-MAML on 24 training tasks combining four MRI contrasts - T1, FLAIR, T2 and PD, two under-sampling mask types - Cartesian and Gaussian and three acceleration factors 4x, 5x and 8x. Tasks are encoded in contrast type - mask type - acceleration factor value. FL denotes FLAIR MRI contrast

Task	SGD	MAML	MMAML	KM-MAML
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1C4	36.14 / 0.891	36.45 / 0.900	36.45 / 0.903	**36.75 / 0.915**
T1C5	35.03 / 0.866	35.22 / 0.875	35.14 / 0.876	**35.58 / 0.895**
T1C8	32.69 / 0.822	32.99 / 0.839	32.96 / 0.842	**33.37 / 0.864**
T1G4	40.07 / 0.913	40.59 / 0.922	40.71 / 0.927	**41.02 / 0.935**
T1G5	38.18 / 0.877	38.67 / 0.887	38.70 / 0.894	**39.05 / 0.906**
T1G8	35.03 / 0.805	35.33 / 0.820	35.51 / 0.841	**35.96 / 0.859**
FLC4	34.12 / 0.879	**34.40** / 0.885	34.23 / 0.882	**34.40 / 0.889**
FLC5	32.87 / 0.849	**33.18** / 0.860	32.96 / 0.854	33.16 / **0.864**
FLC8	30.83 / 0.813	31.02 / 0.828	30.89 / 0.824	**31.20 / 0.841**
FLG4	39.25 / 0.911	**39.42** / 0.907	39.23 / 0.908	39.30 / **0.911**
FLG5	36.95 / 0.875	**37.19** / 0.873	36.97 / 0.876	37.15 / **0.881**
FLG8	33.40 / 0.802	33.73 / 0.811	33.56 / 0.820	**33.93 / 0.832**
PDC4	32.76 / 0.877	33.06 / 0.882	32.96 / 0.879	**33.24 / 0.888**
PDC5	31.17 / 0.835	31.60 / **0.845**	31.39 / 0.840	**31.66 / 0.845**
PDC8	29.07 / 0.803	29.48 / 0.814	29.32 / 0.804	**29.64 / 0.819**
PDG4	37.78 / 0.908	37.98 / 0.911	37.81 / 0.908	**38.03 / 0.914**
PDG5	35.14 / 0.857	**35.39 / 0.863**	35.21 / 0.857	35.37 / **0.863**
PDG8	31.27 / 0.783	**31.72 / 0.791**	31.51 / 0.780	31.69 / 0.789
T2C4	32.01 / 0.873	32.43 / 0.880	32.26 / 0.875	**32.55 / 0.884**
T2C5	30.58 / 0.825	30.96 / 0.834	30.74 / 0.825	31.07 / **0.839**
T2C8	28.24 / 0.787	28.56 / 0.797	28.45 / 0.791	**28.71 / 0.803**
T2G4	37.27 / 0.910	37.48 / 0.912	37.21 / 0.906	**37.59 / 0.916**
T2G5	34.83 / 0.860	35.06 / 0.863	34.82 / 0.855	**35.24 / 0.870**
T2G8	31.06 / 0.779	31.50 / 0.788	31.22 / 0.774	**31.61 / 0.795**

PSNR / SSIM   GT inset   27.65 / 0.743   33.17 / 0.921   33.83 / 0.938   33.30 / 0.924   34.08 / 0.945

Figure A.20: Qualitative Comparison of the reconstruction performance for T2 MRI. From the left: the ground truth (GT) image with the region of interest (ROI), GT inset, ZF image, joint training, MAML, MMAML, and KM-MAML. The yellow arrows and the residual images indicate that KM-MAML can recover repeating patterns much better than other methods.

48

Table A.11: Quantitative comparison of on-the-fly adaptation performance of SGD, MAML, MMAML, and KM-MAML to 24 unseen tasks with deviated acquisition settings

Task	SGD	MAML	MMAML	KM-MAML
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1C6	31.33 / 0.837	31.20 / 0.837	31.23 / 0.832	**31.42 / 0.842**
T1C7	32.65 / 0.846	32.60 / 0.849	32.50 / 0.843	**32.79 / 0.851**
T1C8	30.63 / 0.833	30.61 / 0.834	30.60 / 0.830	**30.87 / 0.842**
T1C9	30.07 / 0.824	30.16 / 0.826	30.08 / 0.822	**30.32 / 0.834**
T1G6	**34.72** / 0.841	34.57 / 0.837	34.34 / 0.832	34.70 / **0.842**
T1G7	**32.44** / 0.794	32.37 / 0.792	32.12 / 0.788	32.41 / **0.797**
T1G8	31.69 / 0.785	31.68 / 0.786	31.28 / 0.774	**31.69 / 0.791**
T1C9	30.98 / 0.779	30.96 / 0.780	30.68 / 0.773	**30.89 / 0.785**
T2C6	32.00 / 0.872	31.94 / 0.871	31.97 / 0.875	**32.16 / 0.877**
T2C7	31.68 / 0.848	31.68 / 0.850	**31.83 / 0.857**	31.73 / **0.857**
T2C8	30.63 / 0.833	30.67 / 0.833	**30.88 / 0.844**	30.76 / 0.842
T2C9	30.47 / 0.854	30.46 / 0.855	**30.66 / 0.862**	30.6 / 0.860
T2G6	35.20 / 0.866	34.72 / 0.855	35.10 / 0.869	**35.27 / 0.870**
T2G7	33.98 / 0.851	33.50 / 0.839	33.82 / 0.855	**34.18 / 0.855**
T2G8	33.37 / 0.837	33.00 / 0.827	33.23 / 0.838	**33.57 / 0.845**
T2G9	32.31 / 0.819	31.78 / 0.804	32.06 / 0.820	**32.41 / 0.823**
PDC6	31.70 / 0.862	31.77 / 0.864	31.70 / 0.867	**31.95 / 0.871**
PDC7	30.45 / 0.843	30.83 / 0.852	30.86 / 0.855	**31.11 / 0.861**
PDC8	31.30 / 0.853	31.43 / 0.857	31.53 / 0.862	**31.85 / 0.868**
PDC9	28.67 / 0.806	28.93 / 0.814	28.81 / 0.812	**29.20 / 0.827**
PDG6	33.80 / 0.851	33.51 / 0.846	33.77 / 0.854	**34.23 / 0.864**
PDG7	32.98 / 0.838	32.71 / 0.831	32.72 / 0.836	**33.26 / 0.851**
PDG8	32.51 / 0.831	32.27 / 0.828	32.52 / 0.833	**32.97 / 0.845**
PDG9	31.46 / 0.803	31.26 / 0.802	31.28 / 0.805	**31.88 / 0.824**

PSNR / SSIM   GT inset   27.65 / 0.743   33.17 / 0.921   33.83 / 0.938   33.30 / 0.924   34.08 / 0.945

Figure A.21: Qualitative Comparison of the reconstruction performance for PD MRI. From the left: the ground truth (GT) image with the region of interest (ROI), GT inset, ZF image, joint training, MAML, MMAML, and KM-MAML. The yellow arrows and the residual images indicate that KM-MAML can recover details better than other methods.

Table A.12: Quantitative comparison of SGD, MAML, MMAML, and KM-MAML in fine-tuning to 10 gradient steps from the meta-initializations to unseen contrasts.

| Task | SGD | MAML | adapt base n/w | | adapt modulation n/w | |
| | | | MMAML | KM-MAML | MMAML | KM-MAML |
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1C4	34.91 / 0.872	**34.94 / 0.875**	34.73 / 0.869	34.90 / 0.874	34.71 / 0.869	34.90 / 0.874
T1C5	33.77 / 0.872	**33.94 / 0.876**	33.48 / 0.867	33.62 / 0.874	33.44 / 0.866	33.61 / 0.873
T1C6	31.33 / 0.837	31.21 / 0.837	31.30 / 0.837	**31.42 / 0.844**	31.25 / 0.835	**31.41 / 0.844**
T1C7	32.67 / 0.847	32.62 / **0.850**	32.47 / 0.843	**32.75 / 0.850**	32.39 / 0.840	**32.73** / 0.849
T1C8	30.64 / 0.834	30.62 / 0.835	30.62 / 0.832	**30.82 / 0.842**	30.62 / 0.832	**30.82 / 0.842**
T1C9	30.08 / 0.824	30.18 / 0.827	30.05 / 0.821	**30.24 / 0.832**	30.05 / 0.821	**30.24 / 0.832**
T1G4	37.87 / 0.892	37.82 / 0.893	37.90 / **0.897**	**37.91** / 0.895	37.84 / 0.895	37.88 / 0.895
T1G5	35.88 / 0.852	35.86 / 0.854	35.75 / 0.850	**35.92 / 0.856**	35.75 / 0.852	**35.92 / 0.856**
T1G6	34.72 / 0.841	34.57 / 0.837	34.41 / 0.834	**34.69 / 0.842**	34.26 / 0.831	**34.65 / 0.841**
T1G7	32.45 / 0.794	32.38 / 0.792	32.15 / 0.789	**32.38 / 0.796**	32.11 / 0.788	**32.37 / 0.796**
T1G8	31.72 / 0.786	**31.72 / 0.788**	31.24 / 0.772	31.56 / 0.786	31.18 / 0.770	31.56 / 0.785
T1G9	**30.99** / 0.779	**30.99** / 0.781	30.83 / 0.780	30.87 / **0.785**	30.70 / 0.775	30.84 / **0.784**
PDC4	33.64 / 0.897	33.71 / 0.899	33.68 / 0.900	**33.98 / 0.904**	33.66 / 0.899	**33.98 / 0.904**
PDC5	32.22 / 0.869	32.54 / 0.876	32.39 / 0.875	**32.71 / 0.882**	32.37 / 0.875	**32.70 / 0.882**
PDC6	31.71 / 0.862	31.77 / 0.864	31.76 / 0.868	**32.00 / 0.872**	31.74 / 0.867	**31.99 / 0.872**
PDC7	30.49 / 0.844	30.85 / 0.852	30.90 / 0.856	**31.15 / 0.861**	30.89 / 0.856	**31.15 / 0.861**
PDC8	31.30 / 0.853	31.43 / 0.857	31.57 / 0.863	**31.89 / 0.868**	31.54 / 0.863	**31.88 / 0.868**
PDC9	28.72 / 0.807	28.95 / 0.815	28.86 / 0.813	**29.20 / 0.827**	28.83 / 0.812	**29.20** / 0.826
PDG4	38.74 / 0.925	38.91 / 0.928	39.06 / 0.931	**39.34 / 0.935**	38.93 / 0.929	**39.30** / 0.934
PDG5	36.46 / 0.894	36.35 / 0.895	36.50 / 0.900	**36.97 / 0.905**	36.50 / 0.900	**36.97 / 0.905**
PDG6	33.84 / 0.852	33.57 / 0.848	33.87 / 0.858	**34.26 / 0.865**	33.81 / 0.856	**34.25** / 0.864
PDG7	32.99 / 0.838	32.72 / 0.832	32.85 / 0.840	**33.29 / 0.852**	32.72 / 0.836	**33.27** / 0.851
PDG8	32.53 / 0.832	32.30 / 0.829	32.48 / 0.832	**32.98 / 0.846**	32.48 / 0.832	**32.96** / 0.845
PDG9	31.50 / 0.804	31.32 / 0.804	31.38 / 0.809	**31.91 / 0.824**	31.30 / 0.805	**31.90 / 0.824**

Table A.13: Quantitative comparison of MAML, MMAML and KM-MAML for five-layer CNN. Meta-training is done on two contrasts (two modes) consisting of T1 and FLAIR contrasts.

| Tasks | MAML | MMAML | KM-MAML |
	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
T1C4	35.97 / 0.9132	35.98 / 0.9115	**36.38 / 0.9189**
T1C5	33.81 / 0.8569	33.91 / 0.8568	**34.16 / 0.8660**
T1C8	31.86 / 0.8248	32.33 / 0.8402	**32.38 / 0.8417**
T1G4	40.16 / 0.9357	40.80 / 0.9441	**40.93 / 0.9473**
T1G5	37.96 / 0.8967	38.21 / 0.8964	**38.30 / 0.9024**
T1G8	33.92 / 0.8232	34.24 / 0.8330	**34.56 / 0.8380**
FLC4	33.32 / 0.8827	33.35 / 0.8829	**33.46 / 0.8832**
FLC5	31.15 / 0.8213	31.23 / 0.8230	**31.30 / 0.8253**
FLC8	29.21 / 0.7947	29.69 / **0.8051**	**29.74** / 0.8004
FLG4	37.97 / 0.8972	38.50 / 0.9129	**38.79 / 0.9158**
FLG5	35.77 / 0.8611	35.62 / 0.8607	**35.80 / 0.8643**
FLG8	31.85 / 0.7875	31.82 / 0.7869	**32.20 / 0.7914**

Table A.14: Quantitative comparison of SGD, MAML, MMAML, and KM-MAML for DenseNet. Meta-training is done on four contrasts consisting of T1, FLAIR, T2, and PD contrasts.

Task	SGD PSNR/SSIM	MAML PSNR/SSIM	MMAML PSNR/SSIM	KM-MAML PSNR/SSIM
T1C5	32.55 / 0.8014	32.15 / 0.7868	33.15 / 0.8271	**33.66 / 0.8384**
T1C8	31.14 / 0.7809	30.80 / 0.7594	31.25 / 0.7880	**31.51 / 0.7975**
T1G5	36.61 / 0.8467	35.98 / 0.8279	37.51 / **0.8771**	**37.80** / 0.8763
T1G8	32.59 / 0.7560	32.09 / 0.7316	33.35 / 0.7899	**33.93 / 0.8091**
FLC5	30.80 / 0.7947	31.12 / 0.8073	31.25 / **0.8199**	**31.33** / 0.8149
FLC8	28.99 / 0.7767	29.09 / 0.7734	**29.26 / 0.7881**	29.17 / 0.7762
FLG5	35.64 / 0.8385	35.95 / 0.8488	36.01 / **0.8513**	**36.06** / 0.8483
FLG8	31.49 / 0.7672	31.82 / 0.7757	32.14 / 0.7927	**32.46 / 0.7949**
PDC5	28.90 / 0.7565	29.47 / 0.7865	29.34 / 0.7720	**30.01 / 0.7898**
PDC8	26.08 / 0.7106	26.48 / 0.7276	26.47 / 0.7214	**27.15 / 0.7371**
PDG5	33.16 / 0.7729	33.90 / **0.8105**	33.69 / 0.7958	**34.00** / 0.7998
PDG8	28.40 / 0.6787	29.08 / 0.7154	28.98 / 0.6939	**30.05 / 0.7222**
T2C5	28.68 / 0.7595	29.08 / 0.7780	28.85 / 0.7669	**29.30 / 0.7801**
T2C8	26.23 / 0.7124	26.47 / 0.7248	26.43 / 0.7225	**26.73 / 0.7306**
T2G5	33.26 / 0.7883	33.80 / 0.8113	33.51 / 0.7992	**33.86 / 0.8038**
T2G8	28.83 / 0.6983	29.38 / **0.7239**	29.14 / 0.6986	**29.87** / 0.7203

Figure A.22: Qualitative Results for T2 MRI comparing the target image, ZF image, joint training, MAML, MMAML, and KM-MAML with respect to the on-the-fly adaptation capabilities of the models to multiple contrasts. The images show that the recovery of repeated image patterns in the predictions of KM-MAML is better than other methods.

51

Figure A.23: PSNR and SSIM plots comparing joint training, MAML, MMAML, and KM-MAML with DenseNet as the base network for four MRI contrasts - T1, FLAIR, PD, and T2. FL denotes FLAIR

Algorithm 2 KM-MAML Fine-tuning the base network for test-time tasks
---
Require: Learning rate η_2 and $p(\mathcal{T})$: Multimodal test task distribution
Require: $CE(.)$: Context encoder
Require: Meta-trained model, ω: modulation network weights and θ: base network weights
 1: Sample a test task mini-batch: $\mathcal{T}_{batch}^{test} \sim p(\mathcal{T})$
 2: **for** each test task M in $\mathcal{T}_{batch}^{test}$ **do**
 3:     Sample a mini-batch of test support data: $D_{spt}^{M,test} = \{x_{US,n}^{M,test}, x_{FS,n}\}_{n=1}^{N_{spt}^{test}}$
 4:     **for** $u = 0$ to $U - 1$ **do**
 5:         Context embedding: $\gamma = CE(x_{US}^{M,test})$ ▷ Support data mini-batch input
 6:         α, $\beta = $ KM Hypernetwork$(\gamma; \omega)$
 7:         $W_{M,test} = \beta \otimes_{outer} \alpha$ ▷ Modulation weights of task M
 8:         $\theta_{mod} = \theta \odot W_{M,test}$ ▷ Mode-specific initialization (KM)
 9:         Initialize $\phi_0^{M,test} \leftarrow \theta_{mod}$
10:         $\mathcal{L}_1 = L([\phi_u^{M,test}, \omega], D_{spt}^{M,test})$
11:         $\phi_{u+1}^{M,test} \leftarrow \phi_u^{M,test} - \eta_2 \nabla_{\phi_u^{M,test}} \mathcal{L}_1$ ▷ mode-specific weight updates
12:     **end for**
13: **end for**

Figure A.24: Validation Mean Square Error (MSE) plots comparing the transfer learning performance of DC-DEN, DC-UNet, with the fine-tuning performance of KM-MAML. Plots show that the validation error values at Epochs 1 and 2 of KM-MAML are comparable with Epochs 8 to 10 of DC-UNet.

Table A.15: List of factors influencing the computation overhead in the training process. For meta-learning methods, the task mini-batch size is set to 3 out of a total of 24 tasks. NA stands for 'Not Applicable' in SGD due to single-loop optimization. 'ms' and 'sec' denote milliseconds and seconds, respectively. For SGD, there is only a single loop for processing the mini-batch of samples. For the meta-learning methods, MAML, MMAML, and KM-MAML, there is a task mini-batch and for each task, there are inner and outer loop iterations. Also, we see that KM-MAML is slightly faster than MMAML in terms of epoch time, as the hypernetworks are meta-learning instead of the base network

Overhead type	SGD	MAML	MMAML	KM-MAML
Inner loop optimization	NA	Yes	Yes	Yes
Outer loop optimization	NA	Yes	Yes	Yes
Outer product computation (Rank 1 approximation) of predicted weights	No	No	No	Yes
Kernel modulation (model-based meta-learning)?	No	No	Yes	Yes
Task mini batch?	NA	Yes	Yes	Yes
No. of optimization loops	1	2	2	2
Which network weights are optimized in two loops?	Base network	Base network	Base network	Hypernetworks
Time taken per iteration (approximately)	5ms per iteration	Inner loop: 4ms per task Outer loop: 1.8 sec per task mini-batch	Inner loop: 7 ms per task Outer loop: 3 sec per task mini-batch	Inner loop: 8 ms per task Outer loop: 2.2 sec per task mini-batch
Time taken per epoch in sec (approaximately)	32	39	59	47