MOST: MR reconstruction Optimization for multiple downStream Tasks via continual learning

Hwihun Jeong[1,2], Se Young Chun[1,3], and Jongho Lee[1†]

[1] Department of ECE, Seoul National University, Seoul, Korea
{hwihuni, sychun, jonghoyi}@snu.ac.kr
[2] Brigham and Women's Hospital, Harvard Medical School, Boston, USA
[3] INMC & IPAI, Seoul National University, Seoul, Korea

Abstract. Deep learning-based Magnetic Resonance (MR) reconstruction methods have focused on generating high-quality images but often overlook the impact on downstream tasks (e.g., segmentation) that utilize the reconstructed images. Cascading separately trained reconstruction network and downstream task network has been shown to introduce performance degradation due to error propagation and the domain gaps between training datasets. To mitigate this issue, downstream task-oriented reconstruction optimization has been proposed for a single downstream task. In this work, we extend the optimization to handle multiple downstream tasks that are introduced sequentially via continual learning. The proposed method integrates techniques from replay-based continual learning and image-guided loss to overcome catastrophic forgetting. Comparative experiments demonstrated that our method outperformed a reconstruction network without finetuning, a reconstruction network with naïve finetuning, and conventional continual learning methods. The source code is available at: https://github.com/SNU-LIST/MOST.

Keywords: Continual learning · MRI reconstruction · Downstream task

1 Introduction

Magnetic Resonance Imaging (MRI) has gained considerable prominence in clinical practice. The advent of deep learning has led to innovations in MRI such as disease classification [8, 17, 24, 29], tumor or lesion segmentation [7, 14], acquisition [28], and image processing [33]. In particular, the scan time reduction has been a focus of innovation, and various reconstruction networks have been proposed to produce high-quality images from highly undersampled k-space data [2, 13, 34]. When evaluating reconstructed images, the gold standard is the assessment by radiologists, which is often impractical due to high costs. Consequently, many studies rely on quantitative metrics such as the Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) [23, 34]. Despite the utility, such metrics have limitations in reflecting the evaluation of radiologists. Moreover, when the reconstruction network is connected to a downstream task network (e.g., segmentation network), optimizing these metrics has been shown to

†Corresponding author.

deliver suboptimal outcomes in the downstream task [9, 10]. As a result, efforts have been made to design a "downstream task-oriented" reconstruction network, optimizing it based on the performance of downstream tasks [11, 30–32]. So far, however, these methods are designed for a single downstream task (Fig. 1a). Hence, when multiple downstream tasks exist and are progressively developed, the current solution is to develop multiple reconstruction networks each optimized for individual tasks, which may not be practical.

In this study, we propose an approach for the MR reconstruction Optimization for multiple downStream Tasks (MOST). Simple expansion from single-task to multi-task optimization is not straightforward [27]. While joint multi-task learning would be ideal when a new task arrives, it is often impractical due to the need to store and retrain all previous data, which is costly in terms of computation and storage. This issue is more critical in the medical domain, where long-term data storage is often restricted by privacy and regulatory constraints. Moreover, MRI data is large, often reaching hundreds of megabytes per scan, and general hospitals may perform over 100,000 scans annually. Given the need for multi-scanner and multi-task datasets, storage demands can quickly reach hundreds of terabytes, making data management challenging. Therefore, we consider a realistic scenario where the reconstruction network is sequentially finetuned with pretrained downstream task networks, assuming limited access to the training dataset of previously trained tasks (Fig. 1b).

When sequentially training a reconstruction network for multiple downstream tasks, one may consider naïve finetuning for the tasks. However, this approach can lead to a well-known issue of catastrophic forgetting, where performance on previously optimized tasks degrades substantially [3, 16, 18–20]. To overcome this issue and sustain the performances for the multiple downstream tasks, we propose to adopt the continual learning [4, 6] and tailor it to our approach.

2 Method

2.1 Finetuning of downstream task-oriented MR reconstruction network

Accelerated MRI reconstruction aims to produce high-quality images from undersampled k-space data, making them comparable to the images from fully sampled data. Traditionally, an MR reconstruction model denoted as f_R with parameters θ is trained using the fidelity loss function $\mathcal{L}_R$ such as SSIM loss:

$$\theta^* = \arg\min_{\theta} \mathcal{L}_R \left(f_R(x;\theta), y \right), \tag{1}$$

where x represents an undersampled k-space data, and y represents an aliasing-free label image.

However, when this reconstruction network is used for a downstream task such as segmentation or classification, a simple cascade of independently trained reconstruction network and downstream network may yield suboptimal results. For example, the reconstructed images may have different characteristics from

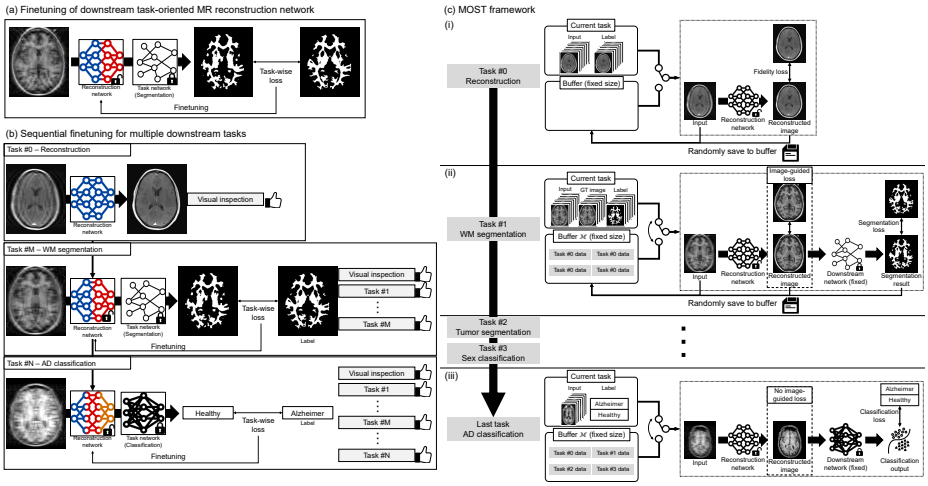

Fig. 1. (a) Combining individually trained reconstruction and downstream networks may result in performance degradation. Downstream task-oriented finetuning can mitigate these challenges. (b) In scenarios with progressively developed downstream tasks, sequential finetuning of the reconstruction network can be an effective solution. However, naïve finetuning can lead to catastrophic forgetting. (c) Overview framework of the MOST

the images trained for the downstream network due to the imperfection of the reconstruction network or domain gaps between the two training datasets. To address this issue, finetuning of the reconstruction network with end-to-end data (i.e., data pairs of undersampled k-space data and downstream task label) using a corresponding loss function is a feasible approach (Fig. 1a) [11, 30–32]:

$$\theta^*_{task} = \arg\min_{\theta} \mathcal{L}_d \left(f_d(f_R(x;\theta)), z \right), \tag{2}$$

where f_d is the downstream task network, $\mathcal{L}_d$ is the corresponding loss function (e.g., cross-entropy for segmentation or classification), and z is the label (e.g., segmentation mask or classification label).

2.2 Sequential finetuning for multiple downstream tasks

When multiple downstream tasks exist, the task-oriented optimization for multiple downstream tasks ($t = 1, 2, 3, \cdots, T$) can be rewritten as follows:

$$\theta^*_{multi-task} = \arg\min_{\theta} \sum_{t} \mathcal{L}_{dt} \left(f_{dt}(f_R(x_t;\theta)), z_t \right), \tag{3}$$

where $\mathcal{L}_{dt}$ is the loss function of the t-th downstream task, and f_{dt} is the t-th downstream task network. The end-to-end dataset of finetuning for the t-th downstream task $D_t = \{x^i_t, z^i_t\}^{n_t}_{i=1}$ includes an undersampled k-space data x and

a downstream task label z with n_t samples. However, such multi-task learning is difficult to be optimized [27] and does not adapt well to real-world applications. In this work, we propose that a single reconstruction network can be sequentially finetuned for multiple downstream tasks (Fig. 1b). Our scenario assumes limited access to previously trained datasets due to privacy regulations and computational costs. Initially, the reconstruction network f_R is trained exclusively for image reconstruction using a loss function ($\mathcal{L}_R$) and a reconstruction dataset (D_R) as written in Eq. 1. Then, we sequentially finetune the reconstruction network for multiple downstream tasks:

$$\theta_t^* = \arg\min_{\theta} \mathcal{L}_{dt}\left(f_{dt}(f_R(x_t;\theta)), z_t\right). \tag{4}$$

Finetuning process exclusively targets the reconstruction network while freezing downstream task networks. When finetuning for the t-th downstream task, we assume that only a small subset of the previous downstream task dataset ($D_p, p < t$) is available.

2.3 MOST

In our proposed approach, MOST, replay-based continual learning is combined with an image-guided (IG) loss to prevent catastrophic forgetting during sequential finetuning for multiple downstream tasks (Fig. 1c). For the replay, a fixed-size buffer ($\mathcal{M}$) is maintained to store a subset of input-output pairs from previous tasks. After finishing each task, we randomly select input-output pairs from that task and add them to the buffer. The buffer size remains constant throughout the finetuning process, with newly added data pairs replacing older ones. Hence, the number of data in the buffer is kept the same across the tasks.

Unlike conventional replay-based continual learning methods [4–6, 25], which generally combine the data of the current and previous tasks in the same mini-batch, our method presents a challenge in creating a combined mini-batch because the input data structure and label type can be different among tasks (e.g., segmentation: 2D input and 2D label vs. classification: 3D input and binary label). To address this challenge, data from the past task in the buffer are used in every K iterations during the finetuning process (Fig. 1c-ii and 1c-iii). The stored data are used in a round-robin fashion across the tasks.

Segmentation often uses high-quality aliasing-free images for label generation, so leveraging that information is reasonable. Therefore, for the downstream task of segmentation, we introduce an additional image-guided loss $\mathcal{L}_{IG}$, to further enhance the finetuning process (Fig. 1c-ii). This loss, which is calculated using a reconstruction loss $\mathcal{L}_R$, computes the difference between the intermediate reconstructed image $f_R(x_t;\theta)$ and the aliasing-free image y_t to enhance the performance of the reconstruction network:

$$\mathcal{L}_{IG} = \mathcal{L}_R(f_R(x_t;\theta), y_t). \tag{5}$$

The reconstructed images are also saved to the buffer to compute the image-guided loss for previous tasks.

3 Experiments and Results

3.1 Dataset and implementation details

We used T_1-weighted images from multiple datasets, including FastMRI [34], OASIS 1 [21], BraTS [22], IXI [1], and ADNI [15] for image reconstruction, white matter (WM) segmentation, tumor segmentation, sex classification, and Alzheimer's disease (AD) classification, respectively. To generate undersampled k-space data, we applied a forward model to the fully sampled image, assuming single-coil acquisition at an acceleration factor of 4. For the segmentation and classification datasets, we divided the dataset into two parts: one for the initial downstream task network training and the other for downstream task-oriented finetuning.

For the reconstruction network, a single coil version end-to-end variational network [13, 34] is deployed. The U-net architecture [26] was used for the network part of the model. We also employed the same U-net architecture for the downstream task of segmentation tasks, and we utilized a CNN for classification tasks. The cross-entropy loss served as a loss function for segmentation and classification tasks. The finetuning process for each downstream task was conducted for a maximum of 5 epochs, and the network with the minimum validation loss was chosen. We utilized buffer data every three iterations ($K = 3$).

We assessed the quality of the reconstructed images using the structural similarity index (SSIM), segmentation results with the Dice similarity coefficient (DICE), and classification performance with the area under the curve (AUC) after completing finetuning for the last downstream task (Last Metric; LM). Additionally, forgetting measures (FM) were computed to quantify how much a model forgets previously learned tasks as it learns new ones. The forgetting measures were calculated by the difference between the best metric and the worst metric of a task obtained during sequential finetuning.

3.2 Performance evaluation

We compared the performance of our MOST approach with a reconstruction network without downstream task-oriented finetuning (Without Finetuning) and a reconstruction network with naïve sequential finetuning (Naïve Finetuning). The buffer size was fixed to 10 subjects. The order of the tasks was WM segmentation, tumor segmentation, sex classification, and finally AD classification.

The qualitative results of the reconstruction and WM segmentation tasks are illustrated in Fig. 2. Without downstream task-oriented finetuning (Fig. 2; third column), the reconstructed images demonstrate good quality for the reconstruction dataset, but exhibit inferior results for the WM segmentation dataset due to domain gaps or error propagation. Naïve Finetuning results in poor quality in both reconstruction and segmentation, indicating catastrophic forgetting (Fig. 2; fourth column). The MOST approach effectively mitigates this issue, producing high-quality results in both tasks (Fig. 2; last column).

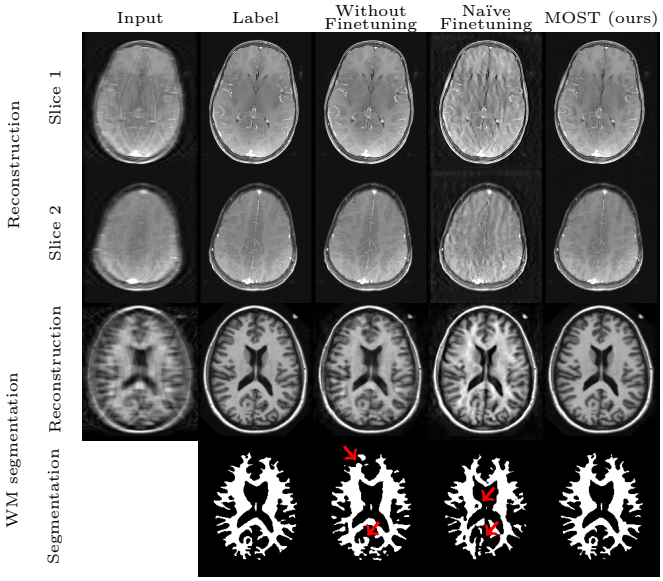

Fig. 2. The results of the reconstruction and WM segmentation before and after the downstream task-oriented finetuning are illustrated.

Table 1. The last metric (LM) and forgetting measure (FM) of SSIM for the reconstruction, DICE for the segmentation, and AUC for the classification are reported.

	Reconstruction SSIM		WM seg DICE		Tumor seg DICE		Sex class AUC		AD class AUC	
	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)
Without Finetuning	0.977	-	0.861	-	0.564	-	0.983	-	0.800	-
Naïve Finetuning	0.866	0.110	0.835	0.098	0.445	0.179	0.976	0.007	0.812	-
MOST (ours)	0.971	**0.006**	**0.938**	**0.002**	**0.642**	**0.003**	0.983	**0.000**	**0.813**	-

Table 1 provides the quantitative metrics. Compared to Without Finetuning, MOST successfully improved performance in all four downstream tasks with a minor decrease in reconstruction quality, demonstrating the effectiveness of finetuning. Compared to Naïve Finetuning, MOST exhibited superior LM and FM, demonstrating its effectiveness in preventing catastrophic forgetting.

3.3 Additional evaluation

Task order experiments were conducted with three different task orders (Table 2a). Compared to Without Finetuning and Naïve Finetuning, MOST exhibited superior LM and FM in most of the tasks and task orders, showing minimal dependency on the task order. When comparing across the orders, starting with segmentation yielded marginal performance benefits, as indicated by a inferior LM and FM in Order 3 when compared to Orders 1 and 2.

Table 2. Result of last metric (LM) and forgetting measures (FM) with three different downstream task orders (a) and comparison result to the other continual learning methods (b) are illustrated.

(a)	Reconstruction SSIM		WM seg DICE		Tumor seg DICE		Sex class AUC		AD class AUC	
	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)
Without Finetuning	0.977	-	0.861	-	0.564	-	0.983	-	0.800	-
	Order 1: Recon → WM seg → Tumor seg → Sex class→ AD class									
Naïve Finetuning	0.866	0.110	0.835	0.098	0.445	0.179	0.976	0.007	0.812	-
MOST (ours)	0.971	**0.006**	**0.938**	**0.002**	**0.642**	**0.003**	**0.983**	**0.000**	**0.813**	-
	Order 2: Recon → Tumor seg → AD class → WM seg → Sex class									
Naïve Finetuning	0.843	0.133	0.930	0.002	0.453	0.181	0.977	-	0.761	0.051
MOST (ours)	0.970	**0.007**	**0.936**	**0.001**	**0.641**	**0.010**	**0.983**	-	**0.804**	**0.012**
	Order 3: Recon → Sex class → AD class → Tumor seg → WM seg									
Naïve Finetuning	0.791	0.186	**0.934**	-	0.421	0.207	0.956	0.031	0.761	0.053
MOST (ours)	0.959	**0.018**	0.920	-	**0.598**	**0.037**	**0.983**	**0.000**	**0.802**	**0.007**

(b)	Reconstruction SSIM		WM seg DICE		Tumor seg DICE		Sex class AUC		AD class AUC	
	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)
Naïve Finetuning	0.866	0.110	0.835	0.098	0.445	0.179	0.976	0.007	0.812	-
EWC [16]	0.910	0.066	0.884	0.048	0.536	0.085	0.976	0.007	0.808	-
LWF [19]	0.937	0.040	0.892	0.026	0.488	0.057	0.893	0.041	0.785	-
ER [6]	0.953	0.024	0.859	0.073	0.479	0.154	0.965	0.018	0.810	-
DER [4]	0.971	0.006	0.929	0.004	0.622	**0.000**	0.981	0.002	0.809	-
MOST (ours)	0.971	0.006	**0.938**	**0.002**	**0.642**	0.003	**0.983**	**0.000**	**0.813**	-

We also **compared the results to conventional continual learning methods** including Elastic Weight Consolidation (EWC) [16], Learning Without Forgetting (LWF) [19], a memory-based method (ER) [6], and another replay-based method (DER) [4]. The results (Table 2b) clearly demonstrated that the MOST approach achieved overall superior performance compared to the other continual learning methods. In contrast, EWC, LWF, ER, and DER showed varying degrees of forgetting and performance degradation. These findings highlight the advantages of MOST over the conventional continual learning methods, mostly due to the fact that the image-guided loss can provide more comprehensive information.

When three **different buffer sizes** (4, 10, and 50 subjects) were evaluated, the results showed consistent performances as shown in Table 3a, indicating that MOST performed well even with the buffer size of 4.

To assess the independent contributions of replay-based continual learning and image-guided loss to the performance of MOST, an **ablation study** was conducted. Three networks were compared: original MOST, MOST without replay-based continual learning, and MOST without image-guided loss. The ablation study results (Table 3b) demonstrated that original MOST outperformed MOST without either replay-based continual learning or image-guided loss. Furthermore, the performance of MOST without image-guided loss confirms that replay-based continual learning alone substantially contributes to performance improvement and catastrophic forgetting reduction.

Table 3. (a) MOST is tested for three different buffer sizes (4, 10, and 50 subjects). (b) Ablation study without replay-based continual learning or image-guided (IG) loss is illustrated.

(a) buffer size	Reconstruction SSIM		WM seg DICE		Tumor seg DICE		Sex class AUC		AD class AUC	
	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)
4	0.970	0.007	0.933	0.007	0.640	0.006	0.982	0.001	0.810	-
10	0.971	0.006	0.938	0.002	0.642	0.003	0.983	0.000	0.813	-
50	0.970	0.007	0.937	0.003	0.640	0.004	0.981	0.003	0.810	-

(b) Replay	IG loss	Reconstruction SSIM		WM seg DICE		Tumor seg DICE		Sex class AUC		AD class AUC	
		LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)	LM(↑)	FM(↓)
X	O	0.950	0.026	0.883	0.056	0.604	0.046	0.976	0.007	**0.817**	-
O	X	0.971	0.006	0.929	0.004	0.622	**0.000**	0.981	0.002	0.809	-
O	O	0.971	0.006	**0.938**	**0.002**	**0.642**	0.003	**0.983**	**0.000**	0.813	-

4 Discussion

In this work, we demonstrated that a single reconstruction network can be sequentially finetuned for multiple downstream tasks. This method successfully generalized recently proposed single downstream task-oriented reconstruction optimization to multiple downstream tasks. In particular, the application of replay-based continual learning along with the image-guided loss successfully prevented catastrophic forgetting, which was observed in naïve finetuning.

MOST demonstrated improved performance in all task orders when compared to that of Naïve Finetuning, but it showed little dependency on task order (Table 2a). In particular, MOST underperformed in Order 3, which may be attributed to the introduction of classification tasks earlier than segmentation tasks. Classification labels contained less detailed information compared to segmentation labels, leading to less effective finetuning.

Our proposed framework is based on the single-coil setting, where the reconstruction model takes single-coil k-space data as the input and produces an alias-free image. The reason for this setting is the limited available end-to-end datasets that include multi-coil k-space data and downstream task labels.

Current continual learning methods mostly target high-level computer vision tasks like classification or segmentation [12], because catastrophic forgetting is rare in regression tasks. In our scenario, although we employ sequential training for the regression network, the cascaded network ultimately performs classifications or segmentations. Hence, catastrophic forgetting occurs in the downstream tasks. If the reconstruction network is merely finetuned with the original reconstruction task for different domains (e.g., images with different hardware or parameters), the effect of catastrophic forgetting may decrease, and continual learning may not be required.

In our approach, we manually trained the downstream task networks instead of adopting off-the-shelf state-of-the-art (SOTA) models, resulting in inferior metrics. This approach was necessary because the SOTA model is trained on the entire training dataset, whereas we required the dataset to be split for down-

stream network training and finetuning. Moreover, the BraTS dataset contained multimodal MRI data, but we only used the T_1-weighted images for consistency across downstream tasks. Additionally, most SOTA models included preprocessing steps on the input images (e.g., skull stripping), which are challenging to be integrated into MOST.

5 Conclusion

We introduced that a single reconstruction network can be sequentially finetuned for multiple downstream tasks and demonstrated that replay-based continual learning and image-guided loss can prevent catastrophic forgetting.

Acknowledgments. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2022R1A4A1030579), Korea Health Industry Development Institute (RS-2024-00439677), Korea Basic Science Institute (RS-2024-00435727), Samsung Electronics Co., Ltd (IO201216-08215-01), INMC and IOER at Seoul National University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ixi - information extraction from images `https://brain-development.org/ixi-dataset`
2. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: Model-based deep learning architecture for inverse problems. IEEE TMI **38**(2), 394–405 (2018)
3. Aljundi, R., Babiloni, F., Elhoseiny, M., Rohrbach, M., Tuytelaars, T.: Memory aware synapses: Learning what (not) to forget. In: ECCV (2018)
4. Buzzega, P., Boschini, M., Porrello, A., Abati, D., Calderara, S.: Dark experience for general continual learning: a strong, simple baseline. In: NeurIPS (2020)
5. Cha, H., Lee, J., Shin, J.: Co2l: Contrastive continual learning. In: ICCV (2021)
6. Chaudhry, A., Rohrbach, M., Elhoseiny, M., Ajanthan, T., Dokania, P.K., Torr, P.H., Ranzato, M.: On tiny episodic memories in continual learning. arXiv preprint arXiv:1902.10486 (2019)
7. Chen, L., Bentley, P., Rueckert, D.: Fully automatic acute ischemic lesion segmentation in dwi using convolutional neural networks. Neuroimage Clinical **15**, 633–643 (2017)
8. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Initiative, A.D.N., et al.: Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database. NeuroImage **56**(2), 766–781 (2011)
9. Desai, A.D., Caliva, F., Iriondo, C., Mortazi, A., Jambawalikar, S., Bagci, U., Perslev, M., Igel, C., Dam, E.B., Gaj, S., et al.: The international workshop on osteoarthritis imaging knee mri segmentation challenge: a multi-institute evaluation and analysis framework on a standardized dataset. Radiology: Artificial Intelligence **3**(3), e200078 (2021)

10. Desai, A.D., Schmidt, A.M., Rubin, E.B., Sandino, C.M., Black, M.S., Mazzoli, V., Stevens, K.J., Boutin, R., Ré, C., Gold, G.E., et al.: Skm-tea: A dataset for accelerated mri reconstruction with dense image labels for quantitative clinical evaluation. arXiv preprint arXiv:2203.06823 (2022)
11. Fan, Z., Sun, L., Ding, X., Huang, Y., Cai, C., Paisley, J.: A segmentation-aware deep fusion network for compressed sensing mri. In: ECCV (2018)
12. Hadsell, R., Rao, D., Rusu, A.A., Pascanu, R.: Embracing change: Continual learning in deep neural networks. Trends in cognitive sciences **24**(12), 1028–1040 (2020)
13. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated mri data. Magnetic Resonance in Medicine **79**(6), 3055–3071 (2018)
14. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Medical Image Analysis **35**, 18–31 (2017)
15. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging **27**(4), 685–691 (2008)
16. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. PNAS **114**(13), 3521–3526 (2017)
17. Li, F., Tran, L., Thung, K.H., Ji, S., Shen, D., Li, J.: A robust deep model for improved classification of ad/mci patients. IEEE Journal of Biomedical and Health Informatics **19**(5), 1610–1616 (2015)
18. Li, X., Zhou, Y., Wu, T., Socher, R., Xiong, C.: Learn to grow: A continual structure learning framework for overcoming catastrophic forgetting. In: ICML (2019)
19. Li, Z., Hoiem, D.: Learning without forgetting. IEEE TPAMI **40**(12), 2935–2947 (2017)
20. Loo, N., Swaroop, S., Turner, R.E.: Generalized variational continual learning. In: ICLR (2021)
21. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience **19**(9), 1498–1507 (2007)
22. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE TMI **34**(10), 1993–2024 (2014)
23. Muckley, M.J., Riemenschneider, B., Radmanesh, A., Kim, S., Jeong, G., Ko, J., Jun, Y., Shin, H., Hwang, D., Mostapha, M., et al.: Results of the 2020 fastmri challenge for machine learning mr image reconstruction. IEEE TMI **40**(9), 2306–2317 (2021)
24. Péran, P., Barbagallo, G., Nemmi, F., Sierra, M., Galitzky, M., Traon, A.P.L., Payoux, P., Meissner, W.G., Rascol, O.: Mri supervised and unsupervised classification of parkinson's disease and multiple system atrophy. Movement Disorders **33**(4), 600–608 (2018)
25. Prabhu, A., Torr, P.H., Dokania, P.K.: Gdumb: A simple approach that questions our progress in continual learning. In: ECCV (2020)
26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
27. Sener, O., Koltun, V.: Multi-task learning as multi-objective optimization. In: NeurIPS (2018)

28. Shin, D., Kim, Y., Oh, C., An, H., Park, J., Kim, J., Lee, J.: Deep reinforcement learning-designed radiofrequency waveform in mri. Nature Machine Intelligence **3**(11), 985–994 (2021)
29. Suk, H.I., Lee, S.W., Shen, D., Initiative, A.D.N., et al.: Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. NeuroImage **101**, 569–582 (2014)
30. Sun, L., Fan, Z., Ding, X., Huang, Y., Paisley, J.: Joint cs-mri reconstruction and segmentation with a unified deep network. In: IPMI (2019)
31. Wang, Z., Xia, W., Lu, Z., Huang, Y., Liu, Y., Chen, H., Zhou, J., Zhang, Y.: One network to solve them all: A sequential multi-task joint learning network framework for mr imaging pipeline. In: MICCAI workshop (2021)
32. Wu, Z., Yin, T., Sun, Y., Frost, R., van der Kouwe, A., Dalca, A.V., Bouman, K.L.: Learning task-specific strategies for accelerated mri. IEEE Transactions on Computational Imaging (2024)
33. Yoon, J., Gong, E., Chatnuntawech, I., Bilgic, B., Lee, J., Jung, W., Ko, J., Jung, H., Setsompop, K., Zaharchuk, G., et al.: Quantitative susceptibility mapping using deep neural network: Qsmnet. Neuroimage **179**, 199–206 (2018)
34. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al.: fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839 (2018)