3D Multi-Coil Brain MR k-Space Dataset for Deep Learning-Based Undersampled Reconstruction

Susanne Schmid[1], Natalia Dubljevic[1], Amirmohammad Shamaei[1], Mariana Bento[1,2,3], Letícia Rittner[4], Richard Frayne[2,5,6,7], and Roberto Souza[1]

[1] Department of Electrical and Software Engineering, University of Calgary, Calgary, Canada
[2] Hotchkiss Brain Institute, University of Calgary, Calgary, Canada
[3] Department of Biomedical Engineering, University of Calgary, Calgary, Canada
[4] School of Electrical and Computer Engineering, University of Campinas, São Paulo, Brazil
[5] Calgary Image Processing and Analysis Centre, Foothills Medical Centre, Calgary, Canada
[6] Seaman Family MR Research Centre, Foothills Medical Centre, Calgary, Canada
[7] Department of Radiology and Clinical Neurosciences, University of Calgary, Calgary, Canada
roberto.souza2@ucalgary.ca

Abstract. The development and validation of deep learning methods to reconstruct magnetic resonance (MR) images from undersampled acquisitions (i.e., fast MR) rely on access to raw k-space data. However, public resources remain limited. This paper presents Calgary-Campinas 2.0 (CC 2.0), an extension of a previously published dataset, containing high-resolution 3D multi-coil brain MR k-space data from 256 acquisitions across 144 subjects. Those scans were acquired using 12-channel and 32-channel coil configurations. In addition to the data expansion, the new release provides longitudinal metadata for repeated scans, including same-day, short-term, and longer-term scan pairs. This structure enables evaluation of reconstruction beyond single-scan quality, including generalization across coil configurations, scan-rescan repeatability, and prior-informed longitudinal reconstruction, in which previous scans may guide the reconstruction of follow-up acquisitions. The dataset captures a longitudinal acquisition setting, with data collected over multiple years across scanner software and hardware updates. CC 2.0 provides a benchmark resource for developing and evaluating MR image reconstruction methods under longitudinal, varying but controlled acquisition conditions.

Keywords: MR image reconstruction · raw k-space · undersampled k-space · accelerated MR · multi-coil MR · longitudinal MR

1 Introduction

Magnetic resonance (MR) imaging is a widely used medical imaging technique that provides excellent soft-tissue contrast without ionizing radiation. However, MR acquisitions are inherently slow. This not only increases patient discomfort and reduces scanner throughput, but also increases the overall cost of imaging. The main reason for the long acquisition time is that MR data are collected in k-space, where sufficient sampling is required to satisfy the Nyquist criterion and avoid aliasing artifacts under conventional Fourier reconstruction. Over the last decades, substantial effort has therefore focused on accelerating MR imaging by improving image reconstructions of incomplete k-space measurements. Established approaches include parallel imaging techniques such as sensitivity encoding (SENSE) [6] and generalized autocalibration partially parallel acquisitions (GRAPPA) [2], as well as compressed sensing (CS) [4]. Those methods differ in their underlying reconstruction strategies, but they aim to reconstruct images from undersampled k-space data. More recently, deep-learning-based reconstruction methods have entered that field to tackle reconstructions of incomplete k-space. Several studies demonstrated improved image quality at higher acceleration factors for deep learning-based techniques [3,10].

However, to train and evaluate machine learning-based MR image reconstruction models, access to raw k-space is required, as demonstrated by public benchmarking efforts such as fastMRI [12]. This is particularly important because reconstruction experiments based on image-derived or otherwise preprocessed data can introduce hidden biases and lead to overly optimistic performance estimates [9]. K-space data are not routinely retained after image reconstruction, partly because of their large file sizes and limited integration into standard clinical workflows. As a result, public datasets containing raw MR k-space data remain relatively limited, especially for high-resolution 3D brain acquisitions. Beyond the clinical advantages of 3D acquisition, volumetric MR also provides a richer setting for accelerated reconstruction. Unlike conventional 2D acquisitions, where undersampling is typically applied along a single phase-encoding direction, 3D acquisitions allow undersampling across two phase-encoding dimensions and thereby enable more flexible sampling patterns and potentially higher acceleration factors. Recent work using 3D multi-coil reconstruction illustrates the growing interest in this setting [11].

The original dataset version has received constant interest over the last few years **(> 10,000 downloads)**, and is one of the few public resources that provides access to raw multi-coil 3D brain MR k-space data [1]. In the new release of Calgary-Campinas 2.0 (CC 2.0), we make four main contributions. (1) We made k-space data available for all scans, including both 12-channel and 32-channel array coil acquisitions. In the original challenge release, the held-out test sets were provided only pre-undersampled for leaderboard evaluation; (2) we expanded the dataset from 67 to 256 MR imaging acquisitions obtained from 144 individual participants; (3) we provided longitudinal metadata of subjects who underwent repeated imaging, including same-day repeat, short-term repeat, and long-term repeat; and (4) we standardized the complex-valued k-space data

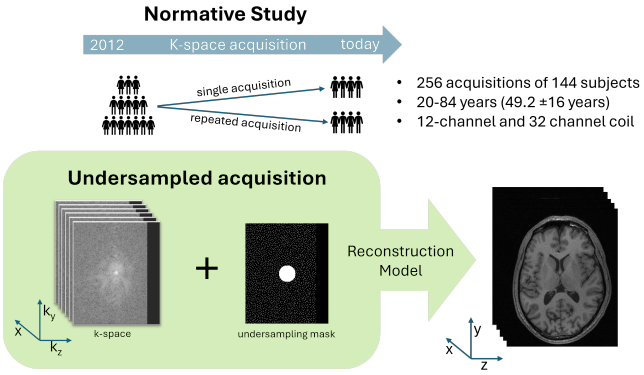


Fig. 1: Overview of the CC 2.0 dataset: 3D T1-weighted brain MR imaging scans were acquired using 12-channel and 32-channel coils of single-subject acquisitions as well as repeated subjects. K-space data is provided along with the metadata, enabling undersampled reconstruction benchmarking, coil generalization experiments, and repeat/longitudinal reconstruction analysis.

by removing an alternating non-anatomical phase modulation along the phase-encoding dimensions. Together, these updates extend the benchmark beyond single-scan reconstruction fidelity and coil generalization toward the evaluation of repeatability, temporal consistency, and robustness in learning-based MR image reconstruction (Figure 1). In the following sections, we will describe the dataset in detail as well as outline benchmark use cases.

2 Dataset Description

CC 2.0 is derived from a real-world longitudinal neuroimaging cohort rather than a purpose-built single-session reconstruction experiment. The data were collected as part of a normative study [5], a multi-year longitudinal study of normal human brain aging, and therefore reflect many of the practical characteristics of long-running MR studies, including repeated participant imaging and routine scanner software and hardware updates over time. K-space data collection started in 2012 and is still ongoing. It was performed on a clinical 3T scanner (Discovery MR750; GE Healthcare, Waukesha, WI, USA). Participants were presumed healthy adults based on cohort screening. MR scans were reviewed for incidental abnormalities, and participants with clinically significant findings outside the expected age-related changes were excluded. Direct identifiers and identifying metadata were removed from the k-space data before release. No defacing or skull-stripping was applied because doing so would require altering the original k-space measurements. The dataset is released under the Creative Commons Attribution NoDerivatives 4.0 International licence (CC BY-ND 4.0 licence). The study and data-sharing procedure were approved by the Research

Table 1: MR image acquisition parameters for the CC 2.0 dataset. Parameters are summarized separately for acquisitions performed with the 12-channel and 32-channel coils. The dataset includes a total of 256 scans, with 144 unique subjects. Six out of the reported eight participants with repeated 32-channel acquisitions had also previously been scanned using the 12-channel coil.

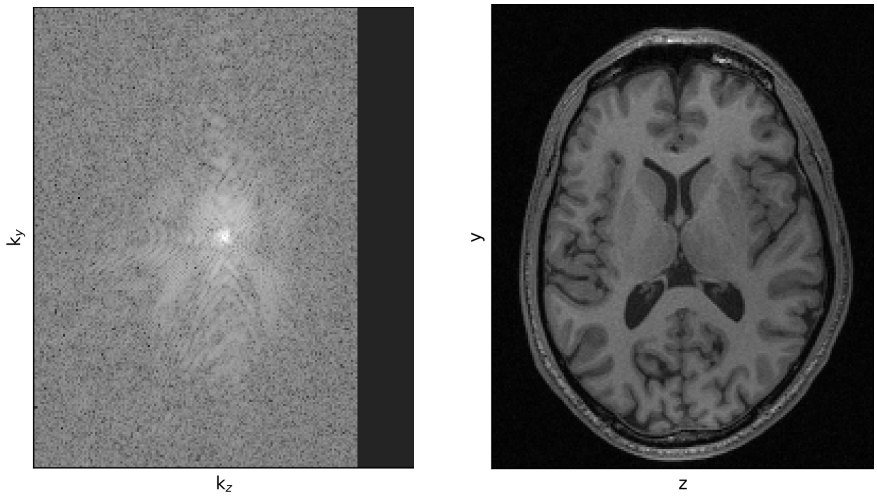
Parameter	12-channel	32-channel
Total number of scans	148	108
Unique subjects	92	58
Repeated subjects (mean scans)	25 (3.24)	8 (7.25)
Scan pairs	59	53
Same day repeat	30	33
Short term repeat (1 - 31 days)	20	13
Long term repeat ($\geq$31 days)	9	7
TR/TE/TI [ms] - flip angle[°]		
6.3/2.6/650 - 8	115 scans	26 scans
7.4/3.1/400 - 10	33 scans	82 scans
Reconstructed FOV [mm^3]	256 x 256 x (170 - 180)	
Acquisition matrix	256 x 218 x (170 - 180)	
Resolution [mm^3]	1 x 1 x 1	
Sequence	Sag 3D-T1w MPRAGE	
Acquisition time [min:s]	5:40	

Ethics Board at the University of Calgary and all participants provided written informed consent.

CC 2.0 contains k-space data of 256 3D T1-weighted brain MR acquisitions acquired from 144 unique participants. The cohort includes 65 males and 79 females, with ages ranging from 20-84 years (49.2 $\pm$ 16 years). One acquisition lasted about 5 min 40 s and was obtained with either 12-channel or 32-channel array coils.

In contrast to the original challenge release, CC 2.0 also exposes repeated-scan metadata, enabling reconstruction studies that examine repeatability. The dataset includes 27 subjects who underwent repeated imaging. These repeated acquisitions were grouped into 112 scan pairs. Out of these scan pairs, 63 were acquired on the same day, 33 scans had short-term intervals within one month (1 - 31 days), and 16 were acquired over longer follow-up intervals. This repeat-scan structure enables evaluation beyond single-scan image quality. It supports assessment of reconstruction repeatability using same-day scan pairs and short-term stability over days to weeks. Long-term scan pairs further enable the development and testing of longitudinal reconstruction methods, where a previous scan may serve as anatomical prior information to improve reconstruction quality or support higher acceleration in follow-up imaging.

Table 1 summarizes the acquisition parameters, including the number of unique subjects and repeated acquisitions for each coil configuration. The data



(a) (b)

Fig. 2: Example of (a) the provided k-space showing the magnitude of the phase-encoding plane with the acquired samples and the zero-filled region; and (b) the corresponding reference reconstruction of the representative reconstructed slice.

were acquired with an acquisition matrix of 256 x 218 x (170 - 180), corresponding to the readout, phase-encoding, and second phase-encoding directions, respectively. The scanner applied a 1D inverse Fourier transformation along one phase-encoding direction. For convenience, the data were converted back into k-space, followed by an inverse Fourier transformation along the readout direction. As a result, the released data is provided as [x, k_y, k_z]. Note that the first phase encoding dimension (k_y) reflects 85 % phase field of view, resulting in 218 acquired samples. In the second phase encoding direction (k_z), the matrix size varied between 170 and 180; however, only approximately 85% of this direction was physically acquired, with the remaining samples being zero-filled (Figure 2). In clinical reconstruction pipelines, such asymmetrical acquired data may be reconstructed using partial Fourier approaches such as homodyne reconstruction techniques.

The other pre-processing step was the removal of an alternating phase modulation along k_y and k_z dimensions. While this modulation did not affect the magnitude images, it produced a checkerboard-like pattern in the phase images. It was therefore removed using the corresponding (± 1) demodulation to be compatible with conventional FFT/iFFT-based approaches.

The dataset is available through the Federal Research Data Repository(FRDR) at doi.org/10.20383/103.01727. Data are provided in HDF5 (.h5) format, with each file containing the provided k-space data in the format of [x, k_y, k_z], with an additional coil dimension for multi-coil acquisitions. Files follow the naming convention: CC_raw_{SubjectID}_{VisitID}. Associated metadata is provided

as *.csv and includes acquisition parameters, coil information, age and sex, recommended train/validation/test splits, and repeat-scan identifiers. A separate *.csv for longitudinal reconstruction studies is provided and contains suggested scan-pairs.

3 Benchmark Use Cases and Evaluation

CC 2.0 is designed to support several complementary benchmark settings for accelerated multi-coil 3D MR image reconstruction. Reconstruction performance can be measured using standard image-quality metrics, including peak signal-to-noise ratio (pSNR), structural similarity index measure (SSIM), and visual information fidelity (VIF). For those metrics, the reconstructions from retrospectively undersampled measurements are compared against the reference reconstructions obtained from the corresponding native data. For standardized benchmark evaluation, the canonical reference reconstruction is obtained by leaving the unacquired k_z locations zero-filled, applying the inverse Fourier transform along the remaining phase-encoding dimensions and combining the individual coil images using square-root sum-of-squares.

When describing reconstruction benchmarks, performance should be reported separately for acquisitions of unique subjects versus repeated-acquisition subsets. Repeated scans represent correlated observations and may disproportionately influence performance metrics when combined with single acquisitions.

As benchmark use cases, we propose three complementary settings:

1. Accelerated 3D multi-coil reconstruction
2. Generalization across coils
3. Repeat-scan and longitudinal reconstruction evaluation

Accelerated 3D multi-coil reconstruction. The provided k-space data can be retrospectively undersampled along the two phase-encoding dimensions to simulate accelerated acquisitions. Retrospective undersampling might be performed using Poisson-disc sampling with a fully sampled central k-space region of 16-pixel radius at different acceleration factors e.g., R=5 and R=10. Undersampling can be applied along a single phase-encoding direction or jointly across both phase-encoding dimensions. Reconstructions from the undersampled measurements can be compared with the reference reconstruction obtained from the fully sampled data.

Generalization across coils. The inclusion of data acquired with both 12-channel and 32-channel coils enables the evaluation of how reconstruction models generalize across different coil configurations. In the proposed data split, only acquisitions obtained with the 12-channel array coil are used for training and validation, while the test set contains separate subsets acquired with both 12-channel and 32-channel array coils (Table 2). This configuration allows evaluation of in-distribution performance on 12-channel test data as well as out-of-distribution generalization to the 32-channel data.

Table 2: Number of scans and unique subjects within the train/validation/test set for 12-channel and 32-channel coil acquisitions. To avoid data leakage, all acquisitions from the same subject are assigned to the same split. Follow-up scans are therefore retained within the same train, validation, or test partition as the corresponding baseline scan.

Data splits	12-channel		32-channel	
	scans	subjects	scans	subjects
Train	54	42	0	0
Validation	18	11	0	0
Test	76	40	108	59

Table 3: Number of scan pairs for same day repeat/short term repeat/long-term repeat and the number of unique subjects, for different coil configurations. Five out of the seven unique subjects in the 32-channel test set are already present in the 12-channel test set.

Data splits repeated scans	12-channel		32-channel	
	scan pairs	subjects	scan pairs	subjects
Train	8/4/1	12	0/0/0	0
Validation	3/4/0	7	0/0/0	0
Test	19/12/8	6	33/13/7	7

Repeat-scan and longitudinal reconstruction evaluation. The metadata for repeated scans (Table 3) enables evaluation of reconstruction behavior beyond single-scan image quality. Same-day scan pairs provide a test-retest setting in which anatomical changes are not expected, while undersampling patterns, noise, and minor differences in subject positioning may still affect the acquired k-space data. These scan pairs can therefore be used to assess whether reconstruction methods produce stable images under near-identical biological conditions.

Short-term scan pairs, acquired between 1 and 31 days apart, provide a complementary setting for evaluating reconstruction robustness. Compared with same-day repeats, these scans may include greater differences in patient positioning, coil loading, and other session-specific acquisition conditions, while major anatomical changes are still expected to be limited. This setting can therefore be used to assess whether reconstruction performance and image-derived measurements remain stable across short-term follow-up imaging.

Long-term follow-up scans enable a different class of benchmark tasks. In this setting, previous scans may be used as anatomical prior information to improve reconstruction of later undersampled follow-up acquisitions, potentially supporting higher acceleration or improved artifact suppression [1,8]. Such prior informed longitudinal reconstruction methods should be evaluated not only for image quality, but also for their ability to preserve true anatomical change.

All of the benchmark settings are proposed to guide evaluation, while baseline reconstruction results are not provided in the present dataset release.

4 Discussion and Limitations

CC 2.0 extends the original multi-coil MR image dataset by substantially increasing the number of available k-space 3D acquisitions, exposing repeated scan metadata and improving k-space preprocessing, including the removal of phase modulation. These additions broaden the scope of the dataset beyond single-scan reconstruction and coil generalization, enabling evaluation of reconstruction stability, and prior-informed longitudinal reconstructions.

A key strength of CC 2.0 is that it is derived from a longitudinal neuroimaging cohort and collected over multiple years. The dataset therefore includes practical sources of variability that are relevant for reconstruction research, including repeated participant imaging, multiple coil configurations, and routine scanner software and hardware updates over time. At the same time, the single-site design provides a relatively controlled setting for evaluating accelerated reconstruction methods. Several limitations should be considered when using the dataset. The acquisition includes partial sampling in k_z direction, which introduces ambiguity in the definition of the reference reconstruction. Although zero-filling provides a simple and reproducible benchmark reference, partial-Fourier reconstruction methods, such as homodyne reconstruction, may more closely reflect clinical reconstruction pipelines. Results using such methods may be reported in supplementary analyses but should be clearly distinguished from those obtained using the canonical benchmark reference (zero-filled baseline).

The dataset is also limited to a single scanner, meaning a consistent vendor, field strength, and imaging site. This controlled acquisition setting is valuable for isolating reconstruction behavior, but it does not support evaluation of generalization across different sites. Similarly, the dataset contains 3D T1-weighted brain MR imaging from a presumed healthy adult lifespan cohort. Although age-related anatomical variability may be present, the dataset does not directly test reconstruction performance in disease-specific populations, nor does it evaluate pulse sequences other than T1-weighted. As acceleration is based on a retrospectively undersampled k-space, the experiments simulate missing k-space measurements but do not capture all effects of prospectively accelerated acquisitions. These may include Eddy current effects, changes in sequence design due to undersampling, among others. As a result, performance under retrospective undersampling should be interpreted as a controlled benchmark rather than a direct substitute for prospective clinical validation [7]. The repeated scan structure introduces additional considerations. Same-day and short-term scan pairs are valuable for evaluating reconstruction repeatability and stability, but differences in subject positioning may impede direct voxel-wise comparisons. Image registrations may be helpful, but might introduce partial-volume effects. In addition, while repeated scans occur in the test set, they are correlated observations and should be reported at the participant level to avoid disproportional influ-

ence in the performance metric. Finally, standard image-quality metrics such as pSNR, SSIM, and VIF provide an objective measure of reconstruction performance, but they do not represent a clinical assessment. Future work may extend CC 2.0 with a standardized baseline experiment and prospective acceleration protocols.

In summary, CC 2.0 substantially extends the original multi-coil MR image reconstruction challenge dataset by increasing the number of available 3D brain MR acquisitions, increasing the amount of scanner-provided k-space data, and exposing longitudinal repeat-scan metadata. The dataset preserves the original benchmark strengths of accelerated reconstruction and cross-coil generalization while enabling new evaluation settings focused on scan-rescan repeatability, short-term reconstruction stability, and prior-informed reconstruction of follow-up imaging. CC 2.0 provides a reproducible resource for developing and evaluating deep learning-based MR image reconstruction methods.

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

References

1. Beauferris, Y., Teuwen, J., Karkalousos, D., Moriakov, N., Caan, M., Yiasemis, G., Rodrigues, L., Lopes, A., Pedrini, H., Rittner, L., Dannecker, M., Studenyak, V., Gröger, F., Vyas, D., Faghih-Roohi, S., Kumar Jethi, A., Chandra Raju, J., Sivaprakasam, M., Lasby, M., Nogovitsyn, N., Loos, W., Frayne, R., Souza, R.: Multi-Coil MRI Reconstruction Challenge—Assessing Brain MRI Reconstruction Models and Their Generalizability to Varying Coil Configurations. Frontiers in Neuroscience **16** (2022). https://doi.org/10.3389/fnins.2022.919186
2. Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A.: Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magnetic Resonance in Medicine **47**(6), 1202–1210 (2002). https://doi.org/10.1002/mrm.10171
3. Knoll, F., Hammernik, K., Zhang, C., Moeller, S., Pock, T., Sodickson, D.K., Akcakaya, M.: Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues. IEEE Signal Processing Magazine **37**(1), 128–140 (2020). https://doi.org/10.1109/MSP.2019.2950640
4. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine **58**(6), 1182–1195 (2007). https://doi.org/10.1002/mrm.21391
5. McCreary, C.R., Salluzzi, M., Andersen, L.B., Gobbi, D., Lauzon, L., Saad, F., Smith, E.E., Frayne, R.: Calgary Normative Study: design of a prospective longitudinal study to characterise potential quantitative MR biomarkers of neurodegeneration over the adult lifespan. BMJ Open **10**(8), e038120 (2020). https://doi.org/10.1136/bmjopen-2020-038120
6. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: SENSE: Sensitivity encoding for fast MRI. Magnetic Resonance in Medicine **42**(5), 952–962 (1999). https://doi.org/10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S

7. Rajput, J.R., Weinmueller, S., Endres, J., Dawood, P., Knoll, F., Maier, A., Zaiss, M.: Death by Retrospective Undersampling - Caveats and Solutions for Learning-Based MRI Reconstructions. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 233–241. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72104-5_23
8. Shamaei, A., Stebner, A., Bosshart, S.L., Ospel, J., Ginde, G., Bento, M., Souza, R.: Enhancing and accelerating brain MRI through deep learning reconstruction using prior subject-specific imaging. Magnetic Resonance Imaging **126**, 110558 (2026). https://doi.org/10.1016/j.mri.2025.110558
9. Shimron, E., Tamir, J.I., Wang, K., Lustig, M.: Implicit data crimes: Machine learning bias arising from misuse of public data. Proceedings of the National Academy of Sciences **119**(13), e2117203119 (2022). https://doi.org/10.1073/pnas.2117203119
10. Souza, R., Bento, M., Nogovitsyn, N., Chung, K.J., Loos, W., Lebel, R.M., Frayne, R.: Dual-domain cascade of U-nets for multi-channel magnetic resonance image reconstruction. Magnetic Resonance Imaging **71**, 140–153 (2020). https://doi.org/10.1016/j.mri.2020.06.002
11. Tanabene, A., G. R., C., Massire, A., Nadar, M., Ciuciu, P.: Benchmarking 3D Multi-Coil NC-PDNet MRI Reconstruction. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (2025). https://doi.org/10.1109/ISBI60581.2025.10980654
12. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., Parente, M., Geras, K.J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M., Vincent, P., Yakubova, N., Pinkerton, J., Wang, D., Owens, E., Zitnick, C.L., Recht, M.P., Sodickson, D.K., Lui, Y.W.: fastMRI: An Open Dataset and Benchmarks for Accelerated MRI (2019). https://doi.org/10.48550/arXiv.1811.08839