Parallel Imaging Reconstruction in Public Datasets Biases Downstream Analysis in Retrospective Sampling Studies

Evan Frenklak[1], Yamin Arefeen[1,2], and Jonathan I Tamir[1]
[1]Chandra Family Department of Electrical and Computer Engineering, UT Austin, Austin, TX, United States, [2]MD Anderson, Houston, TX, United States

Synopsis

Keywords: Data Processing, Image Reconstruction, implicit data crimes

Motivation: We explore the "Implicit Data Crime" of datasets whose subsampled k-space is filled using parallel imaging. These datasets are treated as fully-sampled, but their points derive from (1)prospective sampling, and (2)reconstruction of un-sampled points, creating artificial data correlations given low SNR or high acceleration.

Goal(s): How will downstream tasks, including reconstruction algorithm comparison and optimal trajectory design, be biased by effects of parallel imaging on a prospectively undersampled dataset?

Approach: Comparing reconstruction performance using data that are fully sampled with data that are completed using the SENSE algorithm.

Results: Utilizing parallel imaging filled k-space results in biased downstream perception of algorithm performance.

Impact: This study demonstrates evidence of overly-optimistic bias resulting from the use of k-space filled in with parallel imaging as ground truth data. Researchers should be aware of this possibility and carefully examine the computational pipeline behind datasets they use.

Introduction

Training state-of-the-art MRI reconstruction algorithms, such as those with deeply learned priors[1−3] , depend on large public datasets. Standardized datasets are established tools for evaluation of competing methods' performance[4−6]. Recent work has demonstrated the overly optimistic results of reconstruction algorithms trained on datasets with "Implicit Data Crimes" such as zero-padding and JPEG compression[7].

This abstract reveals an unexplored crime with datasets generated from filling in prospectively subsampled k-space with parallel imaging. Such datasets are presented as fully sampled; that is, with an acceleration factor of 1. Subsequent algorithm development will then apply a retrospective acceleration factor by subsampling the presented k-space and comparing the resultant reconstruction to the image obtained from the original data. The interrelation of these two subsampling steps is an under-explored yet crucial element of modern MR algorithm development pipelines. We present an experimental pipeline to examine the extent of this particular implicit data crime. We compare reconstruction performance based on data that are indeed fully sampled with data that are completed using the SENSE[8] algorithm, demonstrating an initial downstream effect in the form of overly optimistic reconstruction error. Finally, we discuss the patterns observed using various prospective and retrospective accelerations, and we suggest next steps in the study of this phenomenon.

Conceptual Framework

Figure 1(a) describes the conceptual framework of the implicit data crime. A clinical scanner acquires prospectively subsampled data and then produces "Fully-Sampled" k-space through a parallel imaging reconstruction. Datasets will then be generated from this "Fully-Sampled" k-space for subsequent algorithm evaluation. However, as seen in Figure 1(b), true prospectively acquired fully-sampled k-space ($R_{pro} = 1$) differs from "Fully-Sampled" k-space filled in with parallel imaging as evident by the striping in the k-space of $R_{pro} = 2$ and $R_{pro} = 4$, which could lead to overly optimistic bias in the subsequent reconstruction algorithm evaluation. Figure 1(c) shows k-space data from SKM-TEA[6] with similar striping (horizontally) where authors have disclosed that the raw data were GRAPPA-reconstructed at $R_{pro} = 2$.

Figure 2 illustrates our pipeline to evaluate the crime. We begin with fully-sampled k-space that is uniformly subsampled by R_{pro} to get "Prospectively Under-sampled K-space". CG-SENSE reconstructs an image, x_{pro}, and then fake "Fully-sampled K-space' is generated by applying the coil profiles and fourier transform to the x_{pro}. This fake k-space is then subsampled by R_{retro} and reconstructed with CG-sense to produce x_{retro}. We finally compare x_{retro} to x_{pro}, which represents the assumption of using a parallel imaging filled in k-space dataset as ground truth.

Methods

<u>Datasets:</u> We evaluated the crime on three datasets:

- BART[9] generated k-space for an 8-channel Shepp-Logan Phantom with additive Gaussian noise at SNR = $[5, 10, 15, 20]$ dB, averaged over 100 noise instances for each SNR.
- An in-vivo slice from 3D MPRAGE acquired with a 32-channel receive array with IRB approval and informed consent.
- Brain slices selected from 300 subjects in the FastMRI dataset[4,5].

<u>Experimental Details:</u> We performed 2D uniform subsampling with prospective $R_{pro} = [1, 2, 4]$ and retrospective $R_{retro} = [1, 2, 3, 4, 5, 6, 7, 8]$ on

all datasets with a 24-line autocalibration region. All reconstructions used BART's[9] implementation of CG-SENSE and ESPIRIT with a heuristically optimized L2-regularization weightings unique to each instance.

Results

In all figures, $R_{pro} = 1$ (the blue curve) represents the "crimeless" setting.

Figure 3 presents phantom results at various SNR levels. The orange and green curves show normalized root mean squared error (NRMSE) for images reconstructed from prospective subsampling as a comparison point. These curves represent the "Implicit Data Crime", as they yield unrealistically reduced error when compared to the "crimeless" setting (blue curve). Strikingly, yet unsurprising in hindsight, it is possible to achieve zero NRMSE.

Figure 4 demonstrates a similar result on the in-vivo brain acquisition.

Figure 5 expands the in vivo experiment to an average of the experiment over 300 distinct slices from the fastMRI dataset.

Discussion and Next Steps

This study demonstrates evidence of bias resulting from the use of MR datasets presented as "fully sampled" but in fact generated from parallel image reconstruction, that are then used for algorithm evaluation. Researchers should be aware of this possibility and carefully examine the computational pipeline behind datasets they use. Likewise, dataset publishers should detail the techniques used to synthesize unsampled points in k-space.

Extensions of this work will include experiments in which the prospective reconstruction algorithm mismatches the retrospective reconstruction algorithm, such as GRAPPA and recent deep learning techniques for parallel imaging[1–3,10]. In addition, analysis of other downstream tasks might demonstrate other impacts from the implicit data crime. In particular, we hypothesize that k-space sampling trajectory design on parallel imaging filled in datasets will result in suboptimal solutions due to datasets' artificial correlations among candidate k-space points.

Acknowledgements

Evan Frenklak is supported by NSF Award #2313998, subaward #G-1B-022. This work was funded in part by grants from NSF CAREER CCF-2239687 Award, NIH U24EB029240, NSF IFML 2019844, and UT Joint Center for Computational Oncology Postdoctoral Fellowship.

References

1. Jalal, Ajil, et al. "Robust Compressed Sensing MR Imaging with Deep Generative Priors." NeurIPS 2021 Workshop on Deep Learning and Inverse Problems. 2021.

2. Hammernik, Kerstin, et al. "Learning a variational network for reconstruction of accelerated MRI data." Magnetic resonance in medicine 79.6 (2018): 3055-3071.

3. Aggarwal, Hemant K., Merry P. Mani, and Mathews Jacob. "MoDL: Model-based deep learning architecture for inverse problems." IEEE transactions on medical imaging 38.2 (2018): 394-405

4. Knoll, Florian, et al. "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge." Magnetic resonance in medicine 84.6 (2020): 3054-3070.

5. Muckley, Matthew J., et al. "Results of the 2020 fastMRI challenge for machine learning MR image reconstruction." IEEE transactions on medical imaging 40.9 (2021): 2306-2317.

6. Desai, Arjun D., et al. "SKM-TEA: A Dataset for Accelerated MRI Reconstruction with Dense Image Labels for Quantitative Clinical Evaluation." Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2021).

7. Shimron, Efrat, et al. "Implicit data crimes: Machine learning bias arising from misuse of public data." Proceedings of the National Academy of Sciences 119.13 (2022): e2117203119.

8. Pruessmann, Klaas P., et al. "SENSE: sensitivity encoding for fast MRI." Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 42.5 (1999): 952-962.

9. Uecker, Martin, et al. "The BART toolbox for computational magnetic resonance imaging." Proc Intl Soc Magn Reson Med. Vol. 24. 2016.1

10. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acqui- sitions (GRAPPA). Magn Reson Med 2002;47:1202–1210.

Figures

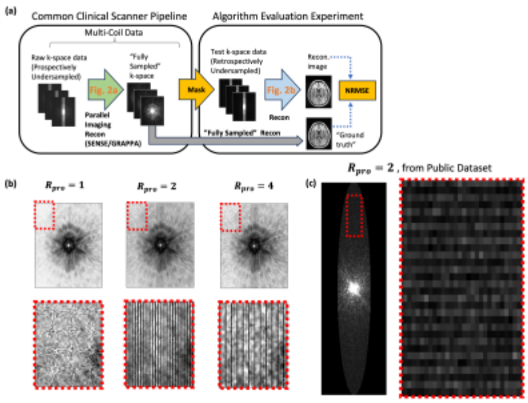

Figure 1. (a) Conceptual framework for the implicit data crime: Prospective undersampling occurs in clinical scanner settings. Parallel imaging recon fills in missing k-space data before it is presented as a "fully sampled" dataset. (b) Stripes are present in the log-magnitude k-space images due to differences between physically sampled points and algorithmically synthesized points. (c) This same patterning occurs in public datasets such as SKM-TEA, again due to parallel imaging reconstruction.

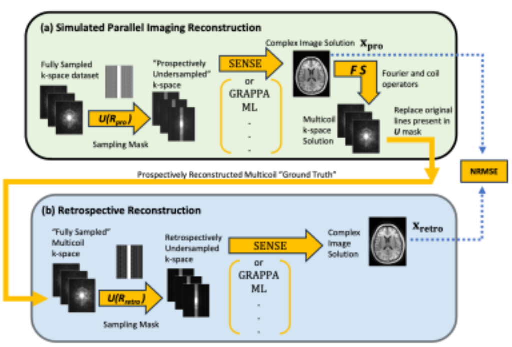

Figure 2. Experimental pipeline for analysis of the implicit data crime. There are two phases: (a) simulated parallel imaging representing the clinical scanner pipeline in Figure 1, and (b) a retrospective parallel imaging reconstruction from a dataset as might be performed in an algorithm evaluation study. We illustrate how each reconstruction might use one of multiple algorithms. Finally (a) and (b) output complex image estimates x_{pro} and x_{retro}, which are compared with a NRMSE metric to illustrate the crime.

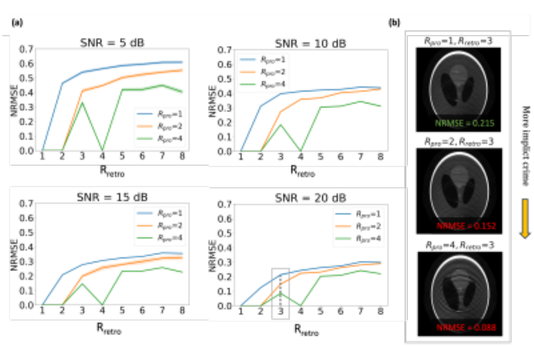

Figure 3. (a) Plots illustrating the implicit data crime on the phantom data at 4 SNR levels where the orange and green curves using parallel imaging filled in k-space as ground truth always yields unrealistically low NRMSE in comparison to the no crime setting (blue curve). Notably, when the two sampling masks match ($R_{pro} = R_{retro}$), we achieve 0 error, as if there were no retrospective acceleration. (b) Example reconstructed images further illustrating the crime.

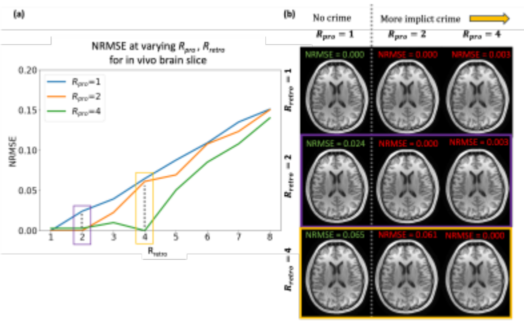

Figure 4. (a) Plots of the implicit data crime on an in-vivo MPRAGE slice where the orange and green curves using parallel imaging filled in k-space as ground truth yields unrealistically low NRMSE. (b) Example reconstructed images. When R_{retro} is a multiple of R_{pro}, the k-space and reconstruction algorithm used to generate the ground truth and image to be evaluated match exactly, resulting in 0% error, a particularly egregious manifestation of the crime.

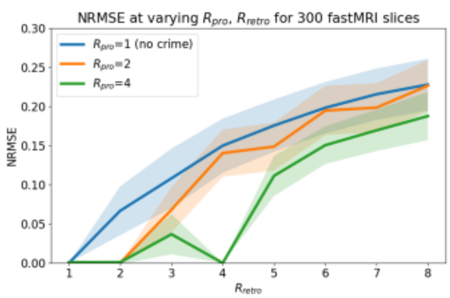

Figure 5. Illustration of the implicit data crime on 300 FastMRI brain samples, where the orange and green curves yield unrealistically reduced error when using a parallel imaging filled in k-space as ground truth in comparison to the "crimeless" setting of the blue curve. Notice how $R_{pro} = 4$ reconstructions (green curve) achieves zero variance at $R_{retro} = 2, 4$ because of consistent alignment between prospective and retrospective sampling masks, while error varies more in the crimeless cases.