Synthesizing Complex-Valued MRI data from Magnitude-Only Images

nikhil deveshwar

Electrical Engineering and Computer Sciences
University of California, Berkeley

August 11, 2023

Acknowledgement

Synthesizing Complex-Valued Multicoil MRI data from Magnitude-only Images

by

Nikhil Deveshwar

A thesis submitted in partial satisfaction of the

requirements for the degree of

Master of Science

in

Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Lustig, Chair
Professor Peder E. Z. Larson

Summer 2023

Synthesizing Complex-Valued MRI data from Magnitude-Only Images

by Nikhil Deveshwar

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Professor Michael Lustig
Research Advisor

(Date)

* * * * * * *

Professor Peder E. Z. Larson
Second Reader

(Date)

Synthesizing Complex-Valued Multicoil MRI data from Magnitude-only Images

Abstract

Synthesizing Complex-Valued Multicoil MRI data from Magnitude-only Images

by

Nikhil Deveshwar

Master of Science in Electrical Engineering and Computer Science

University of California, Berkeley

Professor Michael Lustig, Chair

Despite the proliferation of deep learning techniques for accelerated MRI acquisition and enhanced image reconstruction, construction of large and diverse MRI datasets continues to pose a barrier for effective clinical translation of these technologies. One major challenge is in collecting the MRI raw data (required for image reconstruction) from clinical scanning, as only magnitude images are typically saved and used for clinical assessment and diagnosis. The image phase and multi-channel RF coil information are not retained when magnitude-only images are saved in clinical imaging archives. Additionally, preprocessing used for data in clinical imaging can lead to biased results. While several groups have begun concerted efforts to collect large amounts of MRI raw data, current databases are limited in the diversity of anatomy, pathology, annotations, and acquisition types they contain. To address this, we present a method for synthesizing realistic MR data from magnitude-only data, allowing for the use of diverse data from clinical imaging archives in advanced MRI reconstruction development. Our method uses a conditional GAN based framework to generate synthetic phase images from input magnitude images. We then apply ESPIRiT to derive RF coil sensitivity maps from fully sampled real data to generate multi-coil data. The synthetic data generation method was evaluated by comparing image reconstruction results from training Variational Networks either with real data or synthetic data. We demonstrate that the Variational Network trained on synthetic MRI data from our method, consisting of GAN-derived synthetic phase and multi-coil information, outperformed Variational Networks trained on data with synthetic phase generated using current state of the art methods. Additionally, we demonstrate that the variational networks trained with synthetic k-space data from our method perform comparably to image reconstruction networks trained on undersampled real k-space data.

1 Introduction

Deep learning-based MRI reconstruction methods show promise in faithfully reconstructing MR images from undersampled k-space measurements, but such methods are usually hampered by a lack of paired and diverse training data posing a barrier for effective clinical translation of these technologies. Current deep learning MRI reconstruction techniques use datasets [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] containing paired images and raw k-space MRI data and have enabled major advances in MRI reconstruction methods. However they are limited in several ways. Magnitude images contained in these datasets are sometimes preprocessed which can lead to biased results for MRI reconstruction [14] and are hard to standardize. Furthermore, these publicly available datasets are typically limited in anatomy, acquisition parameters and pathology information. Recent studies have shown that such limitations could sometimes result in hallucinations of structures or artifacts during deep learning based MRI reconstruction[15, 16] limiting the generalization potential of these methods and their clinical use.

There could be significant advantages to leveraging the diversity of existing clinical MRI databases as they contain a range of patient populations, anatomy, pathology, image contrasts, acquisition parameters, and data from different vendors. This would be particularly useful for multi-task networks, e.g.,[17], that perform both image reconstruction and a downstream task such as segmentation or classification. Training on more diverse and representative datasets can also greatly contribute to improving deep learning reconstruction models, especially for rare anatomies and pathologies; this could potentially allow greater clinical adoption.

However, we cannot simply use clinical datasets for MRI reconstruction algorithm development because they typically only contain magnitude images while image phase information is discarded. Furthermore, MRI data is acquired from multi-channel RF coils, but clinical images show a coil-combined image and thus the multi-channel information is lost. MRI phase data is important because it contains information related to contrast from chemical shift, magnetic susceptibility differences, inhomogeneities in the main magnetic field, RF coils used, fat/water separation, tissue interfaces, blood flow, and temperature change [18, 19, 20, 21, 22, 23]. Additionally, recent studies have shown that using complex-valued neural networks which operate on data that includes phase information produce higher-quality reconstructed images. [24, 25]

Thus, the ability to recover or generate image phase from already completed scans could

increase the utility and applicability of deep learning MRI reconstruction methods. While a variety of techniques aim to synthesize different MRI contrasts [26, 27, 28, 29, 30, 31] or parameter maps [32, 33], relatively few techniques exist to synthesize MRI image phase and complex-valued multi-coil data. Recent studies have included methods to generate synthetic image phase by emulating very specific physical models [34], generating sinusoidal phase [35], or have focused on fine tuning training datasets consisting mostly of natural images [36]. To the best of our knowledge, no methods have attempted to broadly synthesize realistic MRI phase maps.

To address this, we present a method for synthesizing realistic MRI data, including image phase and multi-channel information, from magnitude-only images that, for example, are found on clinical imaging archives. Our method leverages recent advances in deep generative modeling [37, 38] to generate synthetic MRI phase images from input MRI magnitude images. Corresponding coil sensitivity maps are derived and then used to generate synthetic multi-channel data. The resulting synthetic multi-coil MRI data, including synthesized image phase, was then evaluated for its ability to be used in image reconstruction tasks by training a variational network [39] and comparing to a network trained on real multi-coil MRI data. Our results show that the proposed method (i) generates realistic looking MR phase maps, (ii) outperforms current methods used to generate synthetic phase data for training reconstruction models and (iii) image reconstruction networks trained on synthetic multi-coil data perform comparably to the same networks trained on real data. Our findings suggest that this framework has the potential to address the limitations that exist in current MRI datasets used for reconstruction tasks where access to raw k-space data is required.

2 Methods

We first start by defining k-space, magnitude, and phase in mathematical terms. We then describe generating synthetic phase images from input magnitude-only images using a conditional generative adversarial network (GAN) framework. Finally we describe the evaluation of the synthetic data quality.

2.1 Preliminaries

The signal acquired from a 2D slice (assuming we can neglect T_2 decay) in the spatial frequency domain, or k-space, can be expressed as:

$$M(k_x, k_y) = \int_X \int_Y m(x, y) e^{-j2\pi(k_x x + k_y y)} dx dy \tag{1}$$

Where $m(x, y)$ is the signal generated at the position (x, y). This is a complex quantity which is equivalent to

$$m(x, y) = m_x(x, y) + j m_y(x, y) \tag{2}$$

where $m_x(x, y)$ is the real component of the signal and $m_y(x, y)$ is the imaginary component of the signal. The goal of MRI reconstruction is to recover $m(x, y)$ from $M(k_x, k_y)$.

The MRI signal, $m(x, y)$, is generated by the rotation of the transverse components of the net magnetization. The signal is complex-valued because it is a measurement of both the x and y components of the net magnetization. The majority of MRI scans are interpreted based on the magnitude of the signal, $|m(x, y)|$, which corresponds to the amplitude of the transverse magnetization. There is also information encoded in the phase (also known as angle) of the signal, $\angle m(x, y)$, which corresponds to the rotation angle of the transverse magnetization. This includes chemical shift, magnetic susceptibility differences, inhomogeneities in the main magnetic field, RF coil profiles, fat/water separation, tissue interfaces, and blood flow.

2.2 Generative Modeling

2.2.1 Neural Network Architecture

The generator is a 16-layer U-Net [40] with skip connections and the discriminator is a 70x70 PatchGAN [38]. In this setup, the discriminator, in a convolutional manner, decides if a patch is real or fake. We use a PatchGAN discriminator to restrict the networks attention to the structure of local image patches. This encourages the discriminator to penalize the structure at the scale of patches rather than the whole image (as in a typical binary classifier) in order to effectively capture high-frequencies in the synthetic image. In a sense, PatchGAN acts as a classifier itself. The main difference is that the output of the PatchGAN is an NxN array where each element signifies whether the corresponding patch in the image is real or fake. We chose a patch size of 70x70 based on results of previous studies [38], which empirically found that this patch size gives the best tradeoff between image sharpness and alleviating artifacts in the generated image. Each generative model was trained with a batch size of 1. We used minibatch stochastic gradient descent (SGD) with the Adam optimizer [41] using

a learning rate of 2e-4 and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Additional implementation details can be found in [38].

2.2.2 Synthetic Phase Generation

The conditional GAN uses a hybrid objective consisting of two loss functions: a conditional adversarial loss function and a regularized L1 distance loss function. In essence, we are training the model to generate high-frequency structures in the synthetic image, and we use the L1 loss to control how much low-frequency structures are present in the image.

$$G^* = arg \min_G \max_D \underbrace{\mathbb{E}_{x,y}[logD(x,y)] + \mathbb{E}_{x,z}[log(1 - D(x,G(x,z)))]}_{\mathcal{L}_{cGAN}} + \lambda \underbrace{\mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1]}_{\mathcal{L}_{L1}} \tag{3}$$

where x is the input magnitude image, y is the generated synthetic phase image that corresponds to the x, and z is the latent vector. In this objective, G tries to minimize the objective while D tries to maximize it. This setup is suitable for our aim because the discriminator is conditioned on the input image x, and we have access to the raw ground-truth data, and thus also to the ground-truth phase data. The network is optimized by alternating between gradient descent steps conducted for optimizing the discriminator and the generator, similarly to the approach as described in the original GAN paper [37].

Specifically, we train a U-Net to predict the phase component from input magnitude-only images. During training, this synthetic phase component is compared to the ground truth phase using a hybrid objective (3). This mixed loss function balances realistic looking phase images via the adversarial loss and encourages less blurring via the L1-norm. Each GAN model was trained for 50 epochs with qualitative analysis of realistic synthetic phase maps being the main stopping criteria. During inference, the trained U-Net is used to generate synthetic phase from previously-unseen magnitude images, resulting in the creation of realistic synthetic phase MRI data.

2.2.3 Multi-coil Data Generation

To generate synthetic multi-coil k-space data we first analytically converted the input magnitude and generated synthetic phase images to real and imaginary components. Sensitivity maps are then generated using the ESPIRiT [42] algorithm on corresponding ground-truth raw data from the training dataset. The resulting sensitivity maps are multiplied with the real and imaginary components to generate multi-coil synthetic k-space data. This resulting complex-valued data can be used in place of ground truth k-space data to train deep learning

based MRI reconstruction networks.

2.3   Dataset

Multi-coil k-space data obtained from the fastMRI [1] dataset was for training the conditional GAN. The dataset consists of raw complex-valued k-space with both magnitude and phase information of brain scans at 1.5 and 3T. The images were acquired with a fast spin echo (FSE) pulse sequence with an echo train length (ETL) of 4. For training we divided the dataset into two datasets with 16-coil and 20-coil acquisitions. Each dataset consisted of T_1-weighted, T_2-weighted, and FLAIR contrast images.

2.4   Experiments

2.4.1   Generative

We trained two generative models: A 16-coil model and a 20-coil model trained on 22691 and 18519 magnitude-only brain images respectively. During training, the U-net generator generates a synthetic phase image and the discriminator compares the generated image to the corresponding ground truth phase image (obtained from fastMRI) in a convolutional patch-wise manner. At inference time, magnitude-only images from the fastMRI test set are run through a forward pass of the trained generative models. In our experiments, this enabled generating 6541 synthetic phase images for the 16-coil model and 5845 synthetic phase images for the 20-coil model.

2.4.2   Evaluation: Physics-based Image Reconstruction

To evaluate the utility of complex-valued multi-coil k-space data synthesized from the generative model, we compared the quality of reconstructed MR images from reconstruction networks trained on ground-truth and synthetic data. Multiple equispaced undersampling masks of acceleration factors $R = \{4, 6, 8, 10\}$ with a center fraction of 0.04 were applied to k-space data to be used for training. Two Variational Networks [39] were then trained for 10 epochs with the 16-coil and 20-coil datasets each. Each Variational Network was trained separately on synthetic and ground truth multi-coil k-space with a 80/10/10 training/validation/test split for a total of four trained reconstruction networks per acceleration factor. Each trained reconstruction model (ground-truth and synthetically trained) was then run on the same ground-truth test set. The quality of reconstructed magnitude images was evaluated using standard quantitative image reconstruction metrics: PSNR, NMSE, SSIM [43]. We decided to use the variational network for evaluation because of its reliance on under-

sampled multi-coil k-space and coil sensitivity maps as inputs into the unrolled network.

All models (generative and reconstruction) were implemented in PyTorch and trained on NVIDIA (Santa Clara, CA) Titan RTX and Quadro 8000 RTX GPUs.
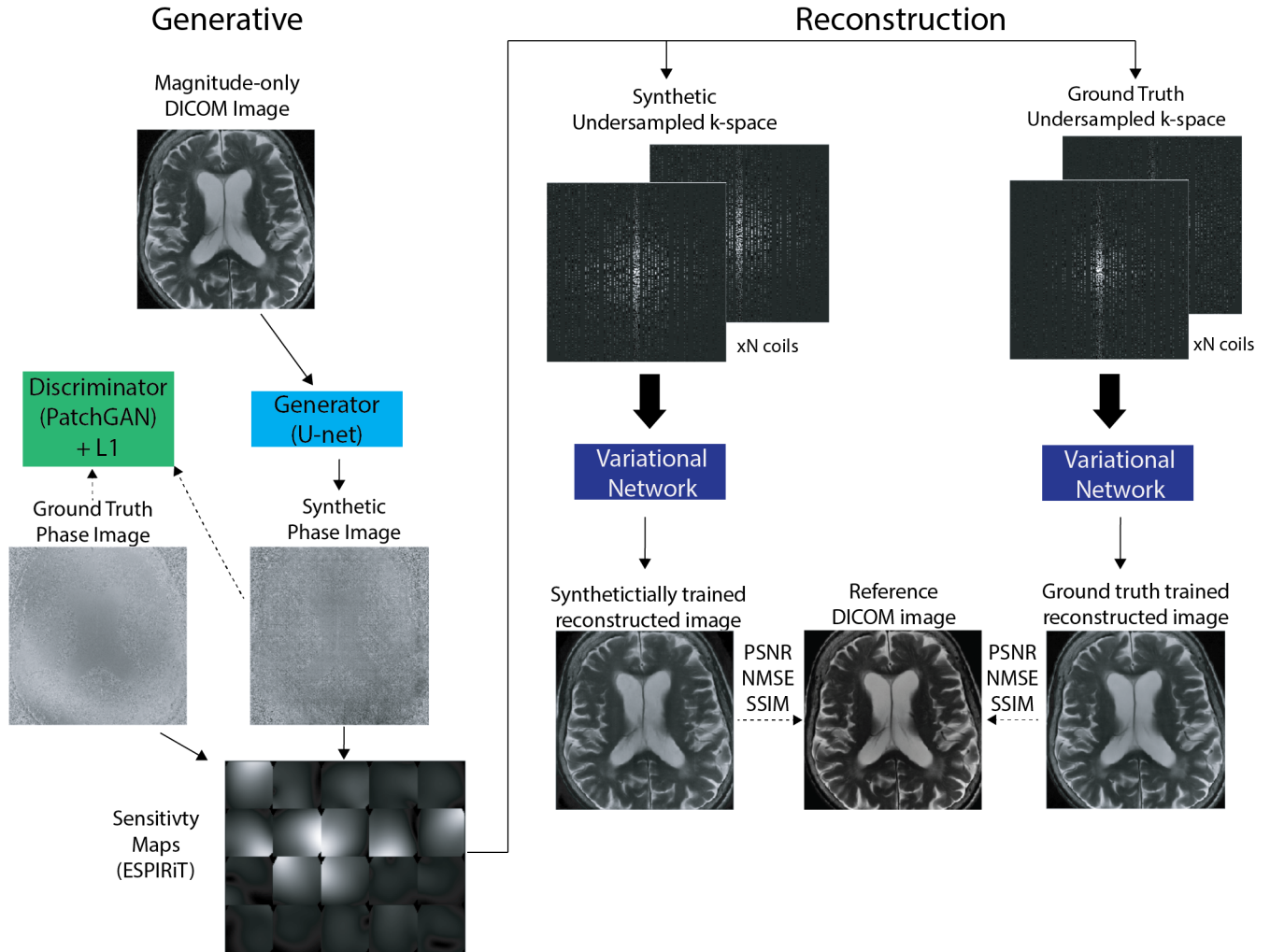

Figure 1: The proposed synthetic raw data generation and image reconstruction pipeline. The generative model takes magnitude images as an input seed and produces plausible synthetic phase images as output, which are trained to match ground truth phase images from the dataset. Synthetic complex-valued data is obtained by combining the input (ground truth) magnitude image and synthetic phase image to yield real and imaginary components. Estimated sensitivity maps calculated with ESPIRiT from the training dataset are then applied to synthetic complex-valued multicoil data to compute multicoil k-space encoded with synthetic phase information. The synthetic raw data was evaluated by training a variational network using undersampled k-space data.

3 Results

Figure 2 shows sample comparisons between synthetic and ground truth phase images. The synthetic phase images show several expected features, including low spatial-frequency components, a noisy background, and tissue phase contrast, for example between the ventricles and adjacent brain tissue. We do not expect the synthetic phase to exactly match the ground truth phase because the MRI phase is not deterministic and can vary based on B_0 homogeneity and RF coil induced phase shifts. In some cases blocking artifacts have appeared, possibly due to the PatchGAN discriminator.
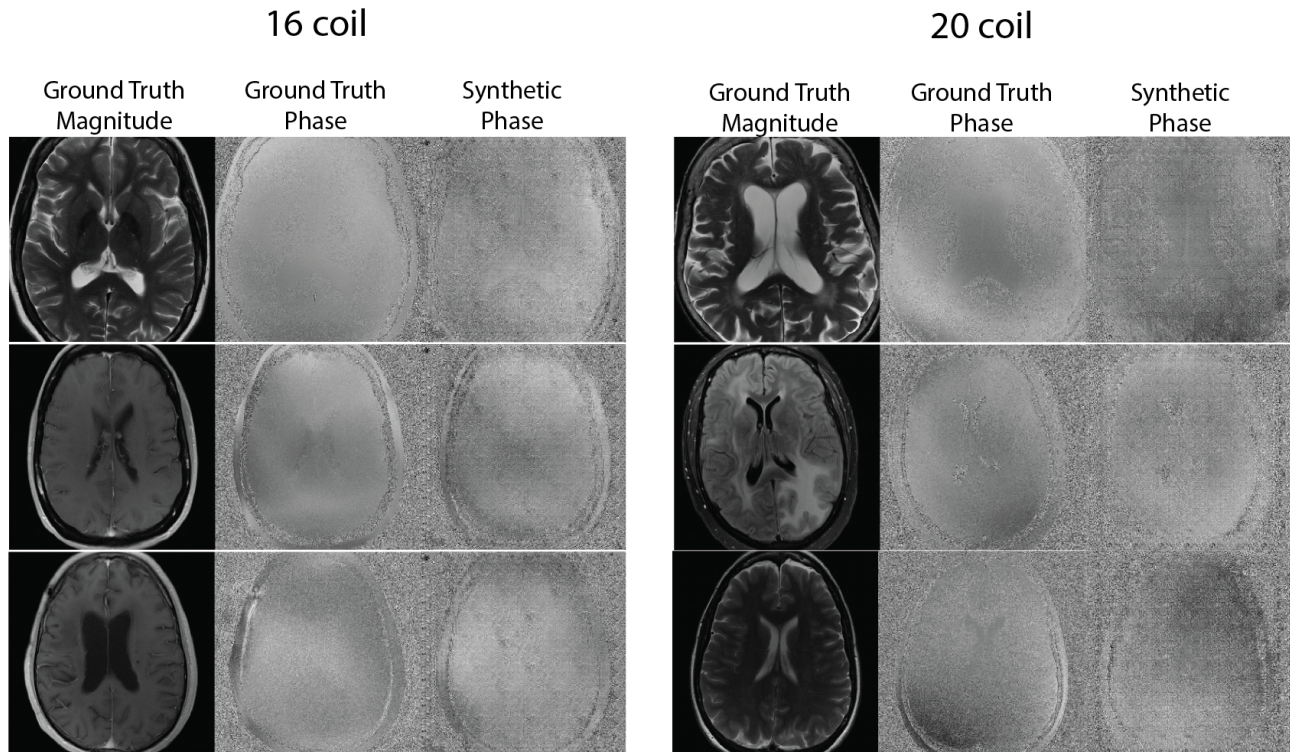
Figure 2: Representative ground truth magnitude, ground truth phase, and synthetic phase images generated from the conditional GAN. Synthetic phase images show expected features, including appropriate noise patterns, low spatial-frequency components and tissue contrast between the ventricles and nearby brain tissue, but exhibit some blocking artifacts possibly from the patchGAN discriminator.

Figures 3 and 4 shows representative images reconstructed with variational network models trained with undersampled ground truth and synthetic k-space data, correspondingly. For $R = 4, 8$ acceleration factors, the reconstructed images trained on synthetic data contain slightly more error structures compared to the images trained on ground-truth data. However, visually there are no obvious artifacts in the reconstructed images in either method. For the $R = 8$ acceleration factor, we can see more errors in high resolution features, possibly due to the lack of high frequency details in the synthetic phase images used to train the reconstruction network.

Figure 5 and the tables in the Supplementary Material compare the effect of different types of phase on variational network reconstruction performance at various acceleration factors. The reconstruction networks were trained on ground truth data, synthetic data (from our proposed method), sinusoidal phase data, data with random phase and data with zero phase. From the plots, variational networks trained on undersampled synthetic data perform comparably to the same networks trained on ground truth undersampled k-space at $R = \{4, 6\}$ as measured by PSNR, NMSE and SSIM. At $R = \{8, 10\}$ acceleration factors, the performance of networks trained on synthetic data dips, especially the SSIM curve, but remains relatively comparable to that of the networks trained on ground truth data. Additionally, the networks trained on synthetic data outperform newtworks trained on sinusoidal phase

Figure 3: Sample image comparisons at 4x and 8x acceleration factors for the 20 coil dataset. The columns compare the zero-filled image, reconstructed image, and error maps generated with 2 variational networks trained on ground truth and synthetic k-space.
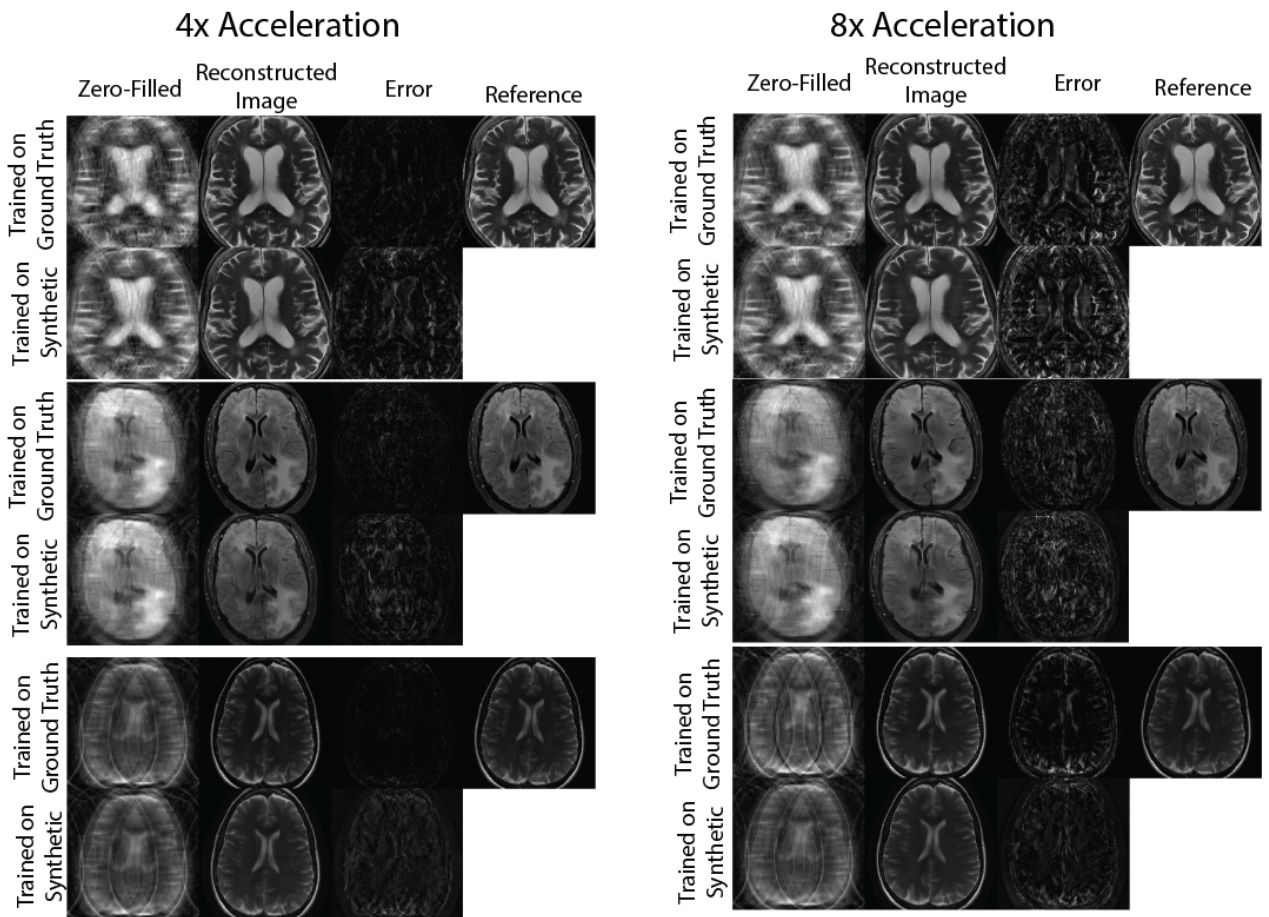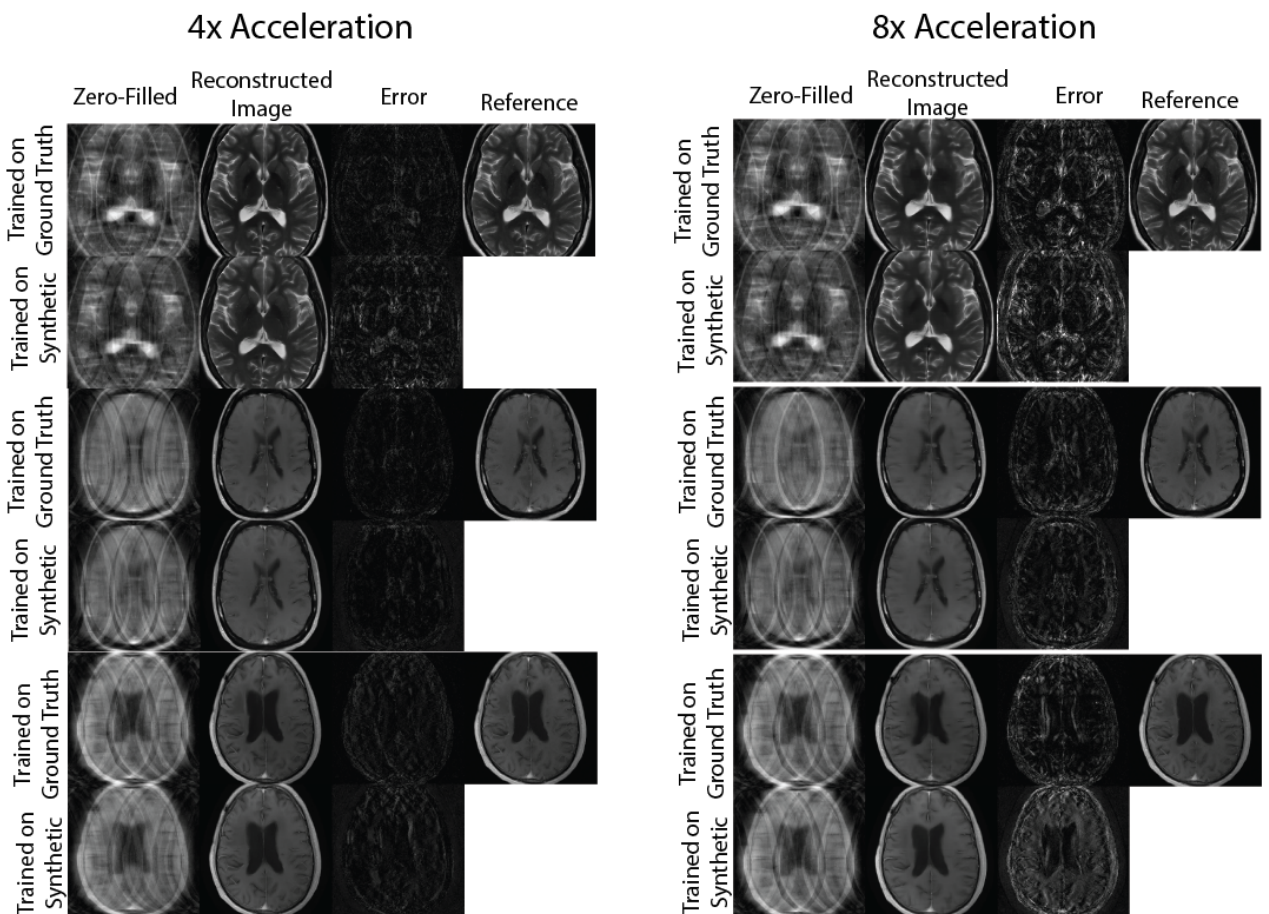

Figure 4: Sample image comparisons at 4x and 8x acceleration factors for the 16 coil dataset. The columns compare the zero-filled image, reconstructed image, and error maps generated with 2 variational networks trained on ground truth and synthetic k-space.

data in all quantitative metrics for the 20-coil dataset. For the 16-coil dataset, similar results were observed for the PSNR and NMSE measurements, while the performance in the SSIM metric was comparable to the sinusoidal phase trained network.
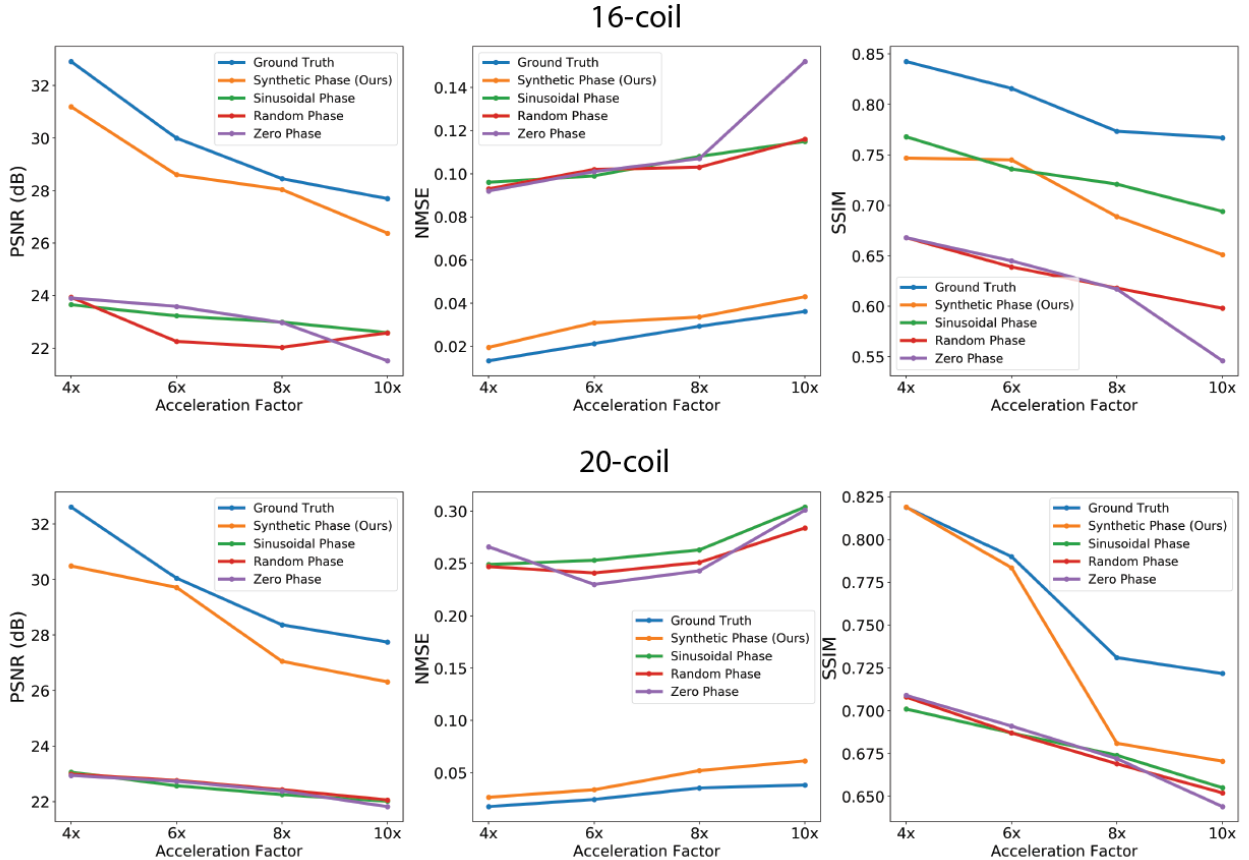

Figure 5: Performance of ground truth-trained and synthetically trained Variational Network reconstruction models at different acceleration factors for 16-coil and 20-coil datasets. At up to 10x acceleration factors, synthetically trained models show comparable performance to ground-truth trained models. This data is also shown in tables in the Supplementary Material.

4 Discussion

There is a massive amount of magnitude-only images as this is what is typically stored in clinical imaging databases (e.g. PACS), which do not usually contain phase and multi-coil information or raw k-space data. This work proposes a framework to generate synthetic multi-coil MRI data from magnitude-only MR images, and evaluates its utility by training a deep learning-based image reconstruction network using the synthesized datasets. The demonstrated framework aims to allow for use of these large imaging databases for developing data-driven methods that require MRI raw data. We chose to evaluate using a variational network image reconstruction model as a proof of concept to demonstrate the effectiveness of the method. We believe a more significant opportunity for such a synthetic data pipeline is to train multi-task networks, e.g. networks that perform both image reconstruction and a downstream task such as image segmentation or classification [17]. In these methods, the

synthetic data pipeline can take advantage of existing clinical images and annotations for the downstream tasks, enabling creation of customized datasets for multi-task machine learning techniques.

Other approaches that generate synthetic MRI training data typically build off natural image datasets. For example in [34], the authors simulated signal voids in MR images by randomly applying masks to natural images to generate synthetic data. Additionally, in [35, 36], the authors used a natural image dataset and a magnitude-only MRI dataset, respectively, and modulated the training images with a sinusoidal phase at random frequency. They demonstrated that training with this synthetic data showed substantially higher levels of aliasing artifacts compared to using real MRI data. The proposed the proposed generative modeling approach shows more realistic image phase maps that include both the low-frequency features which these prior methods aimed to incorporate as well as contrast based on the underlying tissues and anatomy (Fig. 2). Our quantitative results 5 suggest that encoding this tissue phase information (not just low-frequency or sinusoidal phase information) into training data for deep learning models adds more useful information for the network to learn higher quality image reconstructions.

The authors of [36] observe that deviations in SNR, acquisition type, and aliasing patterns between training and test time can result in widely varying image reconstruction quality. With this in mind, future experiments can extend our work to exploit the synthetic data pipeline and large clinical imaging databases to generate custom heterogeneous datasets to train more robust and generalizable image reconstruction models.

In addition to generating synthetic phase maps, a major aim of this work was to generate multi-coil data to increase the clinical relevancy of the technique. We take advantage of the well-established coil sensitivity map algorithm ESPIRiT [42] to estimate coil sensitivities instead of trying to learn them directly. This approach requires running ESPIRiT on prior ground truth data from fastMRI and thus a paired dataset with magnitude and ground truth phase information is still required for this part of the method for image generation.

In previous experiments, we tried to generate multi-coil data by adding a 2 channel real and imaginary component to the output of the conditional GAN. This would result in generated synthetic real and imaginary images for N coils from a single magnitude-only image input. While this approach produced reasonable phase maps and comparable reconstructions for generative models trained on data acquired with a small number of coils (e.g. $N = 4$),

the phase maps resulting from generative models trained on $N = \{16, 20\}$ number of coils suffered from large amounts of structure hallucination and blocking artifacts. We hypothesize that during training, gradients across multiple individual coil images are ill-behaved and thus GAN models trying to generate a large number of coil images have difficulty converging.

The advantage of our proposed technique is that it is coil-agnostic; it can be applied to MR images acquired with any number of coils with the generative model learning a one-to-one mapping from magnitude to phase. This results in more stable training and gradient flow, especially for GANs. It is important to note that we do not expect the synthetic phase maps to be necessarily consistent with the ground truth phase maps. This is because MR phase is not deterministic, and can vary due to tissue composition, scan parameters such as TE, magnetic field homogeneity, and the RF coil configuration and loading. This inconsistency would be problematic for doing any quantification on the synthetic maps themselves. However, consistency with ground truth phase for individual datasets is not required when the synthetic data is used for training, but rather the synthetic phase should be consistent with population-level phase patterns. Nevertheless, enforcing a physics-based consistency between the input magnitude image and output phase image by adding a regularized term in k-space to the training objective could be a useful follow-up experiment to this work. Such a change could result in even more representative phase maps, however GAN stability during training with this new objective remains an open question and would have to be answered empirically. In lieu of this, a score-based generative model could be used for this technique due to their improved training stability compared to GANs [44].

A current limitation of this study is that only fast spin-echo (FSE) images from the fastMRI database were used to train the generative model. The exclusion of gradient-recalled echo (GRE) acquisition data in the training dataset makes the trained generative models and downstream reconstruction models susceptible to distribution shift errors. To address this limitation, future work could include fine-tuning the generative models trained on FSE data with GRE data. Additionally, quantifying the uncertainty in distributions not seen at inference time as proposed in [45] could give insight into how the generative model is synthesizing phase images on a pixel-wise basis.

Finally, evaluation of generative models, especially for synthetic medical imaging data is still an open research direction [46]. While this study used an unrolled image reconstruction network to evaluate the utility of the synthesized complex-valued multi-coil data, other methods, e.g. the Inception Score [47] or FID score [48] could be used to characterize the

distribution of that data. Incorporating a customized implementation of these distance metrics based on medical imaging datasets[49] could be more fruitful in characterizing synthetic phase. This information could also possibly be used to direct generative model training to datasets customized for specific downstream tasks.

5 Conclusion

This work presents a new method for synthesizing realistic, multi-coil MRI data from magnitude-only images that uses a GAN to generate image phase and ESPIRiT-generated coil sensitivity maps. The synthetic data were evaluated by comparing the reconstruction performance of Variational Networks trained on real k-space and synthetic k-space data. Our results suggest that the proposed method for generating synthetic data i) outperforms current state of the art methods for creating synthetic image phase and ii) is adequate for training deep learning MRI reconstruction models at typical acceleration factors (up to 10x), shown by the Variational Networks results. Taken together, our results suggest that image-to-image translation generative adversarial networks are able to generate MRI phase images that are both realistic looking and can also provide good performance when used for training an image reconstruction network. This allows for the possibility of using large, diverse clinical imaging databases that contain magnitude-only images when developing deep learning MRI reconstruction methods.

References

[1] Florian Knoll, Tullie Murrell, Anuroop Sriram, Nafissa Yakubova, Jure Zbontar, Michael Rabbat, Aaron Defazio, Matthew J. Muckley, Daniel K. Sodickson, C. Lawrence Zitnick, and Michael P. Recht. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magnetic Resonance in Medicine*, 84(6):3054–3070, June 2020.

[2] F Ong, S Amin, S Vasanawala, and M Lustig. Mridata.org: An open archive for sharing MRI raw data. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 26, page 1, 2018.

[3] Calgary Campinas Public Dataset. https://sites.google.com/view/calgary-campinas-dataset/mr-reconstruction-challenge?authuser=0, 2020. Accessed: 2021-03-22.

[4] Arjun D Desai, Andrew M Schmidt, Elka B Rubin, Christopher Michael Sandino, Marianne Susan Black, Valentina Mazzoli, Kathryn J Stevens, Robert Boutin, Christopher Re, Garry E Gold, et al. SKM-TEA: A dataset for accelerated MRI reconstruction with dense image labels for quantitative clinical evaluation. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

[5] Yongwan Lim, Asterios Toutios, Yannick Bliesener, Ye Tian, Sajan Goud Lingala, Colin Vaz, Tanner Sorensen, Miran Oh, Sarah Harper, Weiyi Chen, et al. A multispeaker dataset of raw and reconstructed speech production real-time mri video and 3d volumetric images. arXiv preprint arXiv:2102.07896, 2021.

[6] IXI dataset. http://brain-development.org/ixi-dataset/, 2010. Accessed: 2021-03-22.

[7] The human connectome project, University of Souterhn California,. http://www.humanconnectomeproject.org/, 2011. Accessed: 2021-03-22.

[8] AccelMR dataset,. https://accelmrorg.wordpress.com/organizers/#jp-carousel-135/, 2020. Accessed: 2021-05-02.

[9] Oasis dataset,. https://www.oasis-brains.org/, 2007. Accessed: 2021-05-02.

[10] ADNI dataset,. http://adni.loni.usc.edu/data-samples/data-types/, 2004. Accessed: 2021-05-02.

[11] UC Biobank,. https://www.ukbiobank.ac.uk/, 2006. Accessed: 2021-05-02.

[12] The Cancer Imaging Archive. `http://www.cancerimagingarchive.net`, 2010. Accessed: 2021-05-02.

[13] Brain Tumor MRI dataset,. `https://www.kaggle.com/shlezinger/brain-mri-data/`, 2020. Accessed: 2021-05-02.

[14] Efrat Shimron, Jonathan I. Tamir, Ke Wang, and Michael Lustig. Implicit data crimes: Machine learning bias arising from misuse of public data. *Proceedings of the National Academy of Sciences*, 119(13), March 2022.

[15] Kerstin Hammernik, Thomas Küstner, Burhaneddin Yaman, Zhengnan Huang, Daniel Rueckert, Florian Knoll, and Mehmet Akçakaya. Physics-driven deep learning for computational magnetic resonance imaging, 2022.

[16] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne W. Lui, and Florian Knoll. Results of the 2020 fastMRI challenge for machine learning MR image reconstruction. *IEEE Transactions on Medical Imaging*, 40(9):2306–2317, September 2021.

[17] Francesco Caliva', Rutwik Shah, Upasana Upadhyay Bharadwaj, Sharmila Majumdar, Peder Larson, and Valentina Pedoia. Breaking speed limits with simultaneous ultra-fast {mri} reconstruction and tissue segmentation. In *Medical Imaging with Deep Learning*, 2020.

[18] Holger Eggers and Peter Börnert. Chemical shift encoding-based water-fat separation methods. *Journal of Magnetic Resonance Imaging*, 40(2):251–268, January 2014.

[19] Xiang He and Dmitriy A. Yablonskiy. Biophysical mechanisms of phase contrast in gradient echo MRI. *Proceedings of the National Academy of Sciences*, 106(32):13558–13563, August 2009.

[20] Michael Markl, Frandics P. Chan, Marcus T. Alley, Kris L. Wedding, Mary T. Draney, Chris J. Elkins, David W. Parker, Ryan Wicker, Charles A. Taylor, Robert J. Herfkens, and Norbert J. Pelc. Time-resolved three-dimensional phase-contrast MRI. *Journal of Magnetic Resonance Imaging*, 17(4):499–506, March 2003.

[21] Baiyan Zhuang, Arlene Sirajuddin, Shihua Zhao, and Minjie Lu. The role of 4D flow MRI for clinical applications in cardiovascular disease: current status and future per-

spectives. *Quantitative Imaging in Medicine and Surgery*, 11(9):4193–4210, September 2021.

[22] Jeff H. Duyn, Peter van Gelderen, Tie-Qiang Li, Jacco A. de Zwart, Alan P. Koretsky, and Masaki Fukunaga. High-field MRI of brain cortical substructure based on signal phase. *Proceedings of the National Academy of Sciences*, 104(28):11796–11801, July 2007.

[23] Efrat Shimron, William Grissom, and Haim Azhari. Temporal differences (TED) compressed sensing: a method for fast MRgHIFU temperature imaging. *NMR in Biomedicine*, 33(9):e4352, 2020.

[24] Elizabeth Cole, Joseph Cheng, John Pauly, and Shreyas Vasanawala. Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications. *Magnetic Resonance in Medicine*, 86(2):1093–1109, March 2021.

[25] Joshua Bassey, Lijun Qian, and Xianfang Li. A survey of complex-valued neural networks, 2021.

[26] Sooyeon Ji, Dongjin Yang, Jongho Lee, Seung Hong Choi, Hyeonjin Kim, and Koung Mi Kang. Synthetic scpMRI/scp : Technologies and applications in neuroradiology. *Journal of Magnetic Resonance Imaging*, 55(4):1013–1025, November 2020.

[27] Vikas Gulani, Peter Schmitt, Mark A. Griswold, Andrew G. Webb, and Peter M. Jakob. Towards a single-sequence neurologic magnetic resonance imaging examination: Multiple-contrast images from an IR TrueFISP experiment. *Investigative Radiology*, 39(12):767–774, December 2004.

[28] J.B.M. Warntjes, O. Dahlqvist Leinhard, J. West, and P. Lundberg. Rapid magnetic resonance quantification on the brain: Optimization for clinical usage. *Magnetic Resonance in Medicine*, 60(2):320–329, August 2008.

[29] Yongsheng Chen, Saifeng Liu, Yu Wang, Yan Kang, and E. Mark Haacke. STrategically acquired gradient echo (STAGE) imaging, part i: Creating enhanced t1 contrast and standardized susceptibility weighted imaging and quantitative susceptibility mapping. *Magnetic Resonance Imaging*, 46:130–139, February 2018.

[30] Yu Wang, Yongsheng Chen, Dongmei Wu, Ying Wang, Sean K. Sethi, Guang Yang, Haibin Xie, Shuang Xia, and E. Mark Haacke. STrategically acquired gradient echo (STAGE) imaging, part II: Correcting for RF inhomogeneities in estimating t1 and proton density. *Magnetic Resonance Imaging*, 46:140–150, February 2018.

[31] E. Mark Haacke, Yongsheng Chen, David Utriainen, Bo Wu, Yu Wang, Shuang Xia, Naying He, Chunyan Zhang, Xiao Wang, M. Marcella Lagana, Yu Luo, Ali Fatemi, Saifeng Liu, Sara Gharabaghi, Dongmei Wu, Sean K. Sethi, Feng Huang, Taotao Sun, Feifei Qu, Brijesh K. Yadav, Xiaoyue Ma, Yan Bai, Meiyun Wang, Jingliang Cheng, and Fuhua Yan. STrategically acquired gradient echo (STAGE) imaging, part III: Technical advances and clinical applications of a rapid multi-contrast multi-parametric brain imaging method. *Magnetic Resonance Imaging*, 65:15–26, January 2020.

[32] Martina F. Callaghan, Siawoosh Mohammadi, and Nikolaus Weiskopf. Synthetic quantitative MRI through relaxometry modelling. *NMR in Biomedicine*, 29(12):1729–1738, October 2016.

[33] I Blystad, JBM Warntjes, O Smedby, A-M Landtblom, P Lundberg, and E-M Larsson. Synthetic mri of the brain in a clinical setting. *Acta Radiologica*, 53(10):1158–1163, December 2012.

[34] Michael Loecher, Luigi E. Perotti, and Daniel B. Ennis. Using synthetic data generation to train a cardiac motion tag tracking neural network. *Medical Image Analysis*, 74:102223, December 2021.

[35] Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen, and Matthew S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, March 2018.

[36] Florian Knoll, Kerstin Hammernik, Erich Kobler, Thomas Pock, Michael P Recht, and Daniel K Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic Resonance in Medicine*, 81(1):116–128, May 2018.

[37] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[38] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017.

[39] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C. Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational net-

works for accelerated MRI reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 64–73. Springer International Publishing, 2020.

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, 2015.

[41] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[42] Martin Uecker, Peng Lai, Mark J. Murphy, Patrick Virtue, Michael Elad, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, May 2013.

[43] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[44] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*. ACM, August 2022.

[45] Anastasios N Angelopoulos, Amit Pal Kohli, Stephen Bates, Michael Jordan, Jitendra Malik, Thayer Alshaabi, Srigokul Upadhyayula, and Yaniv Romano. Image-to-image regression with distribution-free uncertainty quantification and applications in imaging. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 717–730. PMLR, 17–23 Jul 2022.

[46] Ali Borji. Pros and cons of gan evaluation measures, 2018.

[47] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen, and Xi Chen. Improved techniques for training gans. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[48] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash

equilibrium. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[49] Xueyan Mei, Zelong Liu, Philip M. Robson, Brett Marinelli, Mingqian Huang, Amish Doshi, Adam Jacobi, Chendi Cao, Katherine E. Link, Thomas Yang, Ying Wang, Hayit Greenspan, Timothy Deyer, Zahi A. Fayad, and Yang Yang. RadImageNet: An open radiologic deep learning research dataset for effective transfer learning. *Radiology: Artificial Intelligence*, 4(5), September 2022.

6 Supplementary Material

Table 1: PSNR values for VarNet trained on different types of phase at various acceleration factors for the 16-coil dataset. Bold values indicate the best performing type of phase (not including ground truth)

Phase type	R=4	R=6	R=8	R=10
Ground truth	32.915	29.995	28.4529	27.692
Synthetic Phase (ours)	**31.196**	**28.602**	**28.039**	**26.378**
Sinusoidal Phase	23.652	23.229	22.99	22.590
Random Phase	23.938	22.255	22.028	22.581
Zero Phase	23.911	23.589	22.976	21.518

Table 2: PSNR values for VarNet trained on different types of phase at various acceleration factors for the 20-coil dataset. Bold values indicate the best performing type of phase (not including ground truth)

Phase Type	R=4	R=6	R=8	R=10
Ground Truth	32.605	30.0454	28.367	27.747
Synthetic Phase(ours)	**30.484**	**29.712**	**27.054**	**26.314**
Sinusoidal Phase	23.068	22.576	22.255	22.020
Random Phase	22.994	22.771	22.434	22.066
Zero Phase	2.947	22.74	22.388	21.826

Table 3: NMSE values for VarNet trained on different types of phase at various acceleration factors for the 16-coil dataset. Bold values indicate the best performing type of phase (not including ground truth)

Phase Type	R=4	R=6	R=8	R=10
Ground Truth	0.0133	0.0213	0.029295	0.03615
Synthetic Phase (ours)	**0.0195**	**0.0309**	**0.0336**	**0.043**
Sinusoidal Phase	0.096	0.099	0.108	0.115
Random Phase	0.093	0.102	0.103	0.116
Zero Phase	0.092	0.101	0.107	0.152

Table 4: NMSE values for VarNet trained on different types of phase at various acceleration factors for the 20-coil dataset. Bold values indicate the best performing type of phase (not including ground truth)

Phase Type	R=4	R=6	R=8	R=10
Ground Truth	0.0176	0.0244	0.0354	0.0383
Synthetic Phase (ours)	**0.0265**	**0.0337**	**0.0520**	**0.0613**
Sinusoidal Phase	0.249	0.253	0.263	0.304
Random Phase	0.247	0.2409	0.251	0.284
Zero Phase	0.266	0.230	0.243	0.301

Table 5: SSIM values for VarNet trained on different types of phase at various acceleration factors for the 16-coil dataset. Bold values indicate the best performing type of phase (not including ground truth)

Phase Type	R=4	R=6	R=8	R=10
Ground Truth	0.8425	0.816	0.773	0.767
Synthetic Phase (ours)	0.7468	**0.745**	0.6889	0.6511
Sinusoidal Phase	**0.768**	0.736	**0.721**	**0.694**
Random Phase	0.668	0.639	0.618	0.598
Zero Phase	0.668	0.645	0.617	0.546

Table 6: SSIM values for VarNet trained on different types of phase at various acceleration factors for the 20-coil dataset. Bold values indicate the best performing type of phase (not including ground truth)

Phase Type	R=4	R=6	R=8	R=10
Ground Truth	0.8189	0.790	0.731	0.722
Synthetic Phase (ours)	**0.819**	**0.784**	**0.681**	**0.671**
Sinusoidal Phase	0.701	0.687	0.674	0.655
Random Phase	0.708	0.687	0.669	0.652
Zero Phase	0.709	0.691	0.672	0.644