Learned k-space Partitioning for Optimized self-supervised MRI Reconstruction

Brenden Kadota, Charles Millard, Mark Chiew

***Abstract*— Self-supervised magnetic resonance imaging (MRI) reconstruction methods can train deep learning networks without requiring fully sampled reference data. One such approach, self-supervised via data under-sampling (SSDU), partitions under-sampled k-space into two disjoint sets, with a neural network mapping between them. However, SSDU and its variants rely on heuristic k-space partitioning, which may lead to suboptimal performance and necessitates new partitioning schemes when initial under-sampling patterns change. In this work, we propose a novel approach to learn optimal k-space partitioning by modeling a probability distribution which we use for partitioning. Specifically, we employ the LOUPE framework to learn an optimal partitioning probability distribution. Furthermore, we introduce a weighted dual-domain self-supervised loss function that incorporates both k-space and image-space loss terms. Evaluations on the fastMRI dataset demonstrate that our dual-domain learned partitioning method outperforms existing partitioning strategies and adapts to new sampling patterns without requiring hand-picked partitioning methods.**

***Clinical Relevance*— Typical MRI protocols are under-sampled and reconstructed using parallel imaging. Self-supervised reconstruction can train directly on under-sampled clinical data which can improve diagnostic accuracy and/or reduce scan times.**

I. Introduction

Magnetic Resonance Imaging (MRI) reconstruction is an inverse problem which recovers high-fidelity images from under-sampled Fourier domain (k-space) measurements. Multiple solutions for this inverse problem exist such as parallel imaging (PI) and compressed sensing (CS). PI utilizes the linear independence of the receiver coils for reconstruction [1], [2]. CS casts images into sparse representations to reconstruct incoherently under sampled k-space[3].

Deep learning (DL) reconstruction methods have been proposed to generate high-quality reconstructions by training a network to recover fully sampled k-space from under-sampled measurements. Many schemes have been proposed such as k-space interpolation [4], generative adversarial networks [5], diffusion models [6], [7], and physics informed neural networks [8], [9]. When these networks are trained with large, fully sampled datasets, DL methods can outperform both CS and PI.

More recently in DL reconstruction, end-to-end DL methods have been proposed to jointly optimize k-space under-sampling patterns and reconstruction. Two types of methods for learned under-sampling patterns have emerged: k-space parameterization and probabilistic methods. In k-space parameterization, the k-space locations are parameterized and optimized through gradient descent [10], [11], [12]. Parameterizing k-space no longer constrains points to a cartesian grid resulting in a highly non-convex loss function, requiring additional constraints to the k-space parametrization. Conversely, probabilistic methods learn a sampling probability distribution that is then sampled to generate a mask [13], [14], [15]. As sampling is a discrete operation, these studies rely on relaxations such as a sigmoid [13] or the pass-through relaxation [16]. The gradients of the weights describing the probability distribution are then calculated using the reparameterization trick [17].

In some circumstances, fully sampled k-space is impossible to collect due to factors such as patient motion, signal decay, and time constraints. Additionally, clinical datasets often acquire under-sampled k-space, relying on PI for reconstruction. Self-supervised methods train directly on under-sampled k-space, achieving performance comparable to supervised approaches [18], [19], [20], [21], [22], [23], [24]. Self-supervised reconstruction via data under-sampling (SSDU) [25] is a recent method which partitions under-sampled k-space into two disjoint sets with a network mapping between the two. At inference, the original under-sampled data is passed through the network which generalizes to predict fully sampled k-space. Improvements to SSDU have been made using a dual domain loss function which introduces an image space loss by reconstructing three images from the different sets. [26].

In the original SSDU study, the authors partition k-space based on a 2D gaussian probability distribution which was empirically validated to perform better than random partitioning [25]. Improved partitioning performance was realized when the authors of [27] explained SSDU using the Noiser2Noise framework [28]. By linking SSDU to a self-supervised de-noising processes they showed that partitioning

*Research supported by Natural Sciences and Engineering Research Council (NSERC) of Canada, Canada research chairs program, and Canadian Institutes of Health Research (CIHR).

B. Kadota. and M. Chiew are with the University of Toronto, Toronto, ON, M4N 3M5 Canada and Sunnybrook Research Institute, Toronto, ON, Canada. (Corresponding author contact; phone: 530-220-3241; e-mail: Brenden.kadota@mail.utoronto.ca).

C. Millard is with the Oxford Centre for Functional MRI of the Brain, Oxford, United Kingdom.

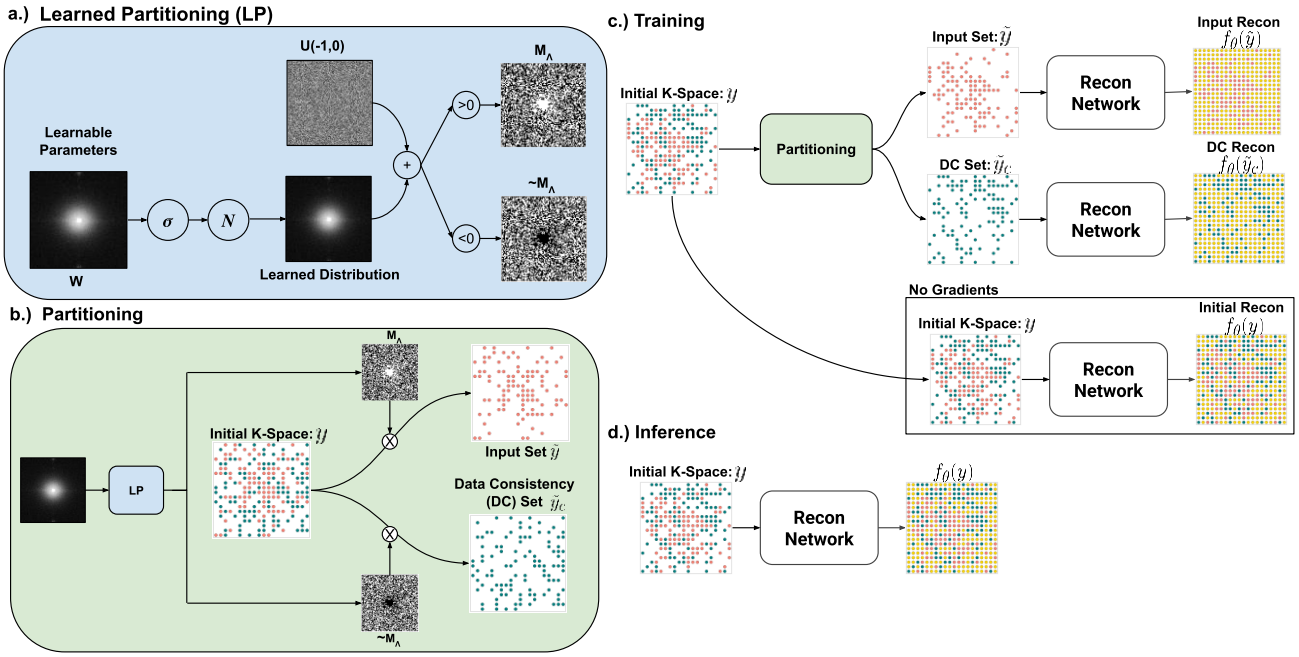

Fig. 1. **a.)** Learned partitioning module for learning optimal partitioning distribution. The distiribution is sampled y adding noise and thresholding **b.)** Partitioning module that partitions undersampled k-space into two sets based on M and ~M. **c.)** Forward pass of reconstruction network during training. Three estimated fully sampled k-spaces from $\tilde{y}$, $\hat{y}$, and y **d.)** At inference only y is reconstructed.

by the same distribution that generated the under-sampled training data improved performance. Further empirical validation to their justification was provided in a study utilizing radial self-supervised sampling based on radial spokes [29].

However, the theoretical justification for SSDU does not provide insight on how to select distribution parameters. This leaves room for potentially improved performance by explicitly learning the partitioning distribution in an end-to-end manner. Notably, one study optimized the initial under-sampling mask for SSDU in and end-to-end manner which improved self-supervised performance [30]. However, to the best of our knowledge, no work has attempted to learn the optimal partitioning of SSDU directly.

Our contributions are summarized as follows:

- We proposed a method to learn the optimal partitioning of k-space for self-supervised learning in an end-to-end manner by learning a probability distribution to partition k-space.
- We empirically test our method on the fastMRI dataset and different initial under-sampling patterns and show that learned partitioning consistently outperforms heuristically chosen masks.
- We conduct ablation studies to empirically justify the dual domain loss function to prevent model collapse.

II. Methods

A. Self-Supervised Learning

The MRI acquisition can be modeled as a forward process defined as:

$$y = M_\Omega A x + n \tag{1}$$

where $x \in \mathbb{C}^n$ is a complex vectorized image, A is the system matrix containing the Fourier transform and coil sensitivities, $y \in \mathbb{C}^n$ is the acquired under-sampled k-space data, and n is random gaussian noise. M_Ω is a diagonal under-sampling matrix drawn from some prior probability distribution. The acceleration factor is determined by $R = \frac{n}{m}$ where m is the number of sampled points in the mask $\sum Tr(M_\Omega)$. Self-supervised learning trains directly on under-sampled k-space y by splitting k-space into two disjoint sets. Given some mask M_Λ, under-sampled k-space can be partitioned into two sets:

$$\tilde{y} = M_\Lambda y \;\; and \;\; \tilde{y}_c = {\sim}M_\Lambda y \tag{2}$$

where $\tilde{y}$ and $\tilde{y}_c$ are disjoint k-space sets and ${\sim}M_\Lambda$ is the complement of M_Λ (Fig 1b). It is important to note that M_Λ partitions *under-sampled k-space* into two sets. The choice of M_Λ is often drawn from a hand-picked probability distribution.

A neural network is trained to predict k-space points in $\tilde{y}_c$ from $\tilde{y}$. This mapping is learned through the loss function:

$$\mathcal{L}(M_\Omega{\sim}M_\Lambda f_\theta(\tilde{y}), \tilde{y}_c) \tag{3}$$

Where f_θ is a neural network parameterized by θ and $\mathcal{L}$ is some chosen loss function. In this work, we employ the $\ell 1$ norm for $\mathcal{L}$ due to its ability to preserve sharper features. As f_θ outputs an estimate for all points in k-space, the masking terms $M_\Omega{\sim}M_\Lambda$ ensure the loss is only calculated where points $\tilde{y}_c$ exist. At inference, the initial un-partitioned data y is passed through the network to get a fully sampled k-space estimate.

B. Learned Partitioning

We propose a method to learn the probability distribution of M_Λ by extending LOUPE [13] for self-supervised learning. We define a learned probability distribution P as:

$$\sigma_t(W) = P \tag{4}$$

where $W \in \mathbb{R}^n$ are learnable sampling weights, σ_t is the element wise sigmoid with a slope controlled by the hyperparameter t: $\sigma_t(W) = \frac{1}{1+e^{-tW}}$. We deviate from LOUPE by keeping the probability distribution unnormalized, allowing the model to learn the acceleration factor for the partitioning distribution.

We sample from the probability distribution P by using the reparameterization trick [17]:

$$M_\Lambda = (P - U) > 0. \tag{5}$$

Here, $U \in \mathbb{R}^n$ is a uniform distribution over [0, 1] sampled independently for each voxel. During gradient calculations, this operation is relaxed by using the passthrough method [16]. During back-propagation, the gradient of the discrete sampling step (4) is replaced by the gradient of a sigmoid function with a slope of 300. A schematic of thew whole process is described in Fig 1a.

C. *Loss*

Our network has three reconstruction outputs during model training. $\tilde{y}$, $\tilde{y_c}$, and y are independently passed through the same reconstruction network to produce three fully sampled k-space estimates: $f_\theta(\tilde{y}), f_\theta(\tilde{y_c})$, and $f_\theta(y)$ (Fig. 1c). To ensure consistency and parameter efficiency, the reconstruction network f_θ shares weights across all three inputs. We choose to disable gradient calculations during the forward pass of $f_\theta(y)$ due to improved running time as we alleviate the need to backpropagate through reconstruction calculations for y (Fig. 1c).

We calculate a dual-domain loss for both multi-coil k-space and image space. Instead of the original self-supervised loss term which predicts $\tilde{y}_c$ from $\tilde{y}$ we add an additional term that predicts the inverse, $\tilde{y}$ from $\tilde{y}_c$. This adds an additional term to the loss function (Fig. 2a):

$$\mathcal{L}_{k_1} = \|M_\Omega \sim M_\Lambda f_\theta(\tilde{y}) - \tilde{y}_c\|_1^1 \tag{6a}$$
$$\mathcal{L}_{k_2} = \|M_\Omega M_\Lambda f_\theta(\tilde{y}_c) - \tilde{y}\|_1^1 \tag{6b}$$

To improve performance, a dual domain loss similar to [26] is introduced by additionally computing an image space loss between all three estimated k-spaces: $f_\theta(\tilde{y}), f_\theta(\tilde{y_c})$, and $f_\theta(y)$.

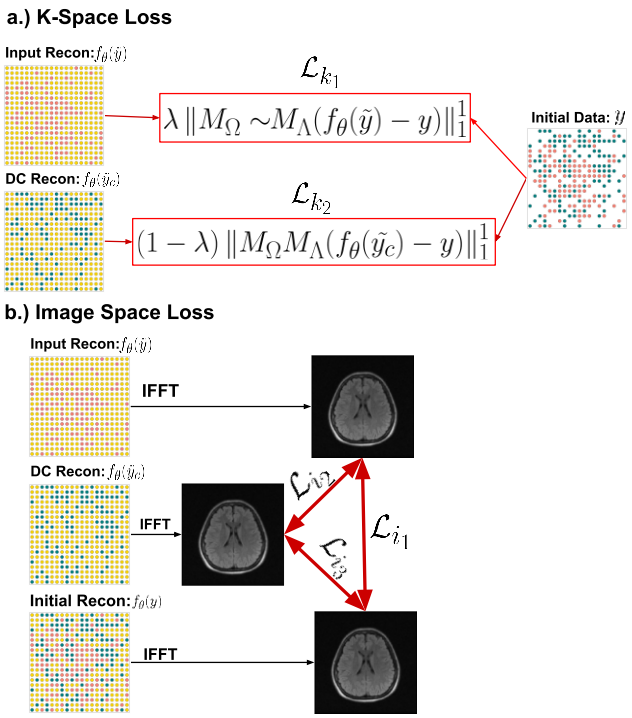

Fig 2. **a.)** Self-supervised k-space loss **b.)** Image space loss. $\mathcal{L}_{i_1}$ is the image loss between the input and DC recon. $\mathcal{L}_{i_2}$ is the image loss between input and initial recon. $\mathcal{L}_{i_3}$ is the image loss between DC and initial recon.

The image space loss function is:

$$\mathcal{L}_{i_1} = SSIM(rss(F^{-1}f_\theta(\tilde{y})),\ rss(F^{-1}f_\theta(y))))$$
$$\mathcal{L}_{i_3} = SSIM(rss(F^{-1}f_\theta(\tilde{y}_c)),\ rss(F^{-1}f_\theta(y))))$$
$$\mathcal{L}_{i_3} = SSIM(rss(F^{-1}f_\theta(\tilde{y}_c)),\ rss(F^{-1}f_\theta(y)))) \tag{7}$$

Where F^{-1} is the inverse Fourier Transform, rss is the root sum of squares operator, and SSIM is the structural similarity index (SSIM) (Fig. 2b).

The final loss function is:

$$\mathcal{L} = \lambda\mathcal{L}_{k_1} + (1-\lambda)\mathcal{L}_{k_2} + \beta_1\mathcal{L}_{i_1} + \beta_2\mathcal{L}_{i_2} + \beta_3\mathcal{L}_{i_3} \tag{8}$$

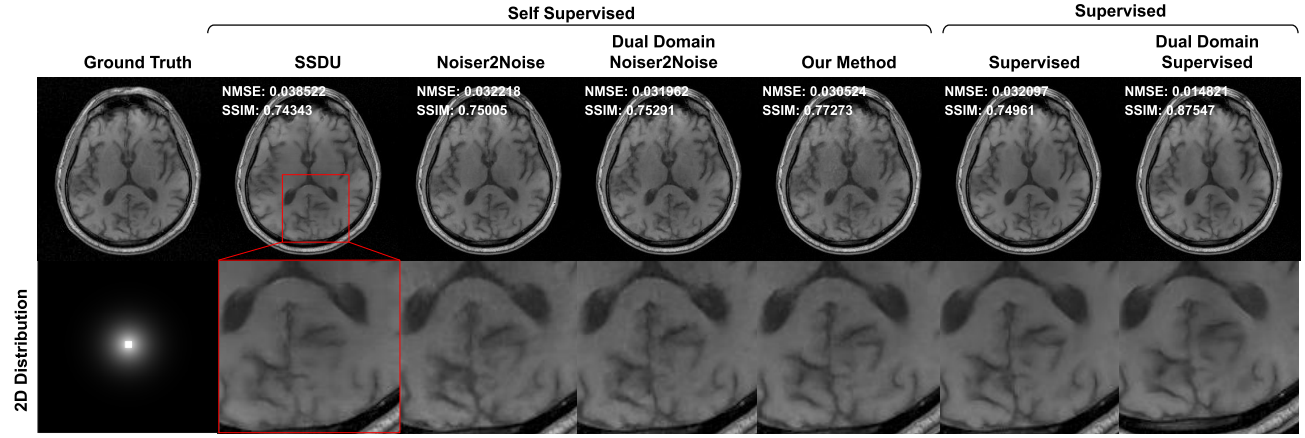

Figure II.1 Reconstructed images from different deep learning methods on the top row. Bottom left is initial undersampling distribution. Bottom row is zoom of red box to better display fine details.

Where β_1, β_2, and β_3 are image space loss scaling factors and λ is a k-space loss scaling factor (Fig. 2). λ is scaled between [0, 1] to balance the importance of the $f_\theta(\tilde{y})$ and the $f_\theta(\tilde{y}_c)$ reconstruction output on the loss. At inferences the network can generalize to fully sampled k-space by passing the original under-sampled k-space to the network $f_\theta(y)$.

D. Implementation Details

We implement all models in PyTorch and train on a multi-instance NVIDIA A100 GPU with 20 GB of memory, utilizing 3/8 of the total compute power. We use the Adam optimizer [31] with a learning rate of 10^{-3}, train for 50 epochs with a batch size of 3. We set the k-space loss scaling factor λ to 0.60 and β_1 to $7.1x10^{-7}$, β_2 to $7x10^{-7}$ and β_3 to $8x10^{-8}$ based on a random hyperparameter sweep. For f_θ we use an end-to-end VarNet [9] which is an unrolled physics informed neural network that estimates coil sensitivities and consists of data consistency and a denoising neural network steps. The center 10x10 region of k-space is left sampled for both $\tilde{y}$ and $\tilde{y}_c$ for coil sensitivity estimation. The VarNet is unrolled for 6 cascades for a reconstruction network with 15M parameters. The learned partitioning model adds an additional 65K parameters.

E. Datasets

We evaluate our method using the fastMRI dataset, a multi-channel raw k-space brain dataset. [32]. We exclude the test data and partition the training and validation sets ourselves. For simplicity, we use 16 channel data and the T1-weigthed images pre and post contrast administration. We split the remaining data volume wise into 145, 18, and 19 volumes (2874, 282, 292 slices) for training, validation, and testing. We preprocess the data by removing alternating lines in the phase-encoding direction and apply zero-padding to a matrix size of [320, 320] to maintain a consistent input size. We set the network outputs for zero-padded regions to zero.

F. Comparative Methods

We train 3 other self-supervised models as comparative methods. All methods use the same VarNet reconstruction network. We train SSDU with gaussian partitioning based on the original SSDU paper which we call SSDU [25]. We also train a SSDU variant that is partitioned with the same distribution as the original sampling distribution M_Ω which we call Noiser2Noise [27]. The last comparative self-supervised network is partitioned using the same strategy as the Noiser2Noise model but is trained with the dual domain loss function.

TABLE I. QUANTITATIVE PERFORMANCE METRICS ON DIFFERENT INITIAL DISTRIBUTIONS

	1D Distribution		Parallel Imaging	
	SSIM	NMSE	SSIM	NMSE
SSDU	0.74092	0.03301	0.73667	0.04487
Noiser2Noise	0.75083	0.03249	0.76182	0.03391
Dual Domain Noiser2Noise	0.75098	0.03174	0.75973	0.03456
Dual Domain Learn Partitioning	**0.76959**	**0.03141**	**0.76437**	**0.03359**
Supervised	0.75513	0.03117	0.75321	0.03151
Dual Domain Supervised	0.87421	0.01655	0.87324	0.01474

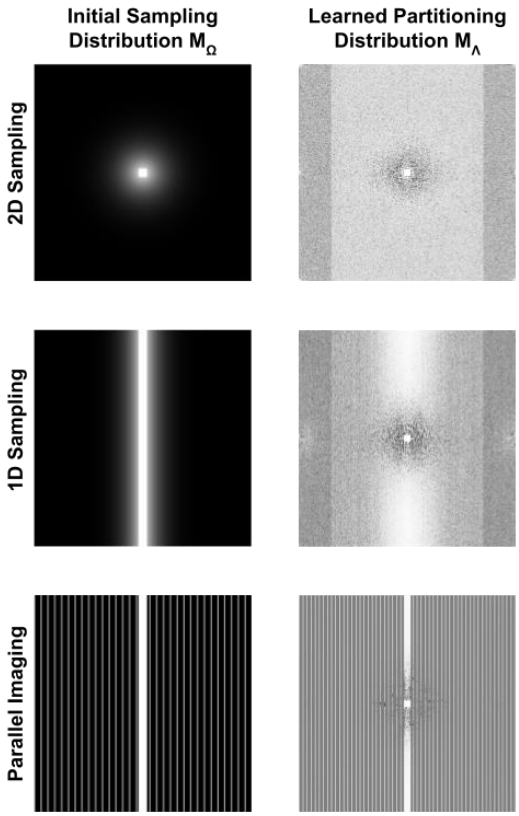

Fig 4. The resulting learned partitioning distributions with the initial undersampling distributions.

We additionally train two supervised reconstruction networks as a baseline. We first trained a supervised VarNet with a k-space only $\ell 1$ loss. We next train a supervised dual domain VarNet with an $\ell 1$k-space loss and SSIM image space loss. The SSIM loss is scaled relative to the $\ell 1$ loss by a factor of 10^{-4}. All comparative methods have the same number of parameters and only differ based on partitioning and loss functions.

Our first experiments explored the qualitative and quantitative effects of the initial sampling masks M_Ω at an acceleration factor of R=6. We tested 3 masks: a 1D and 2D polynomial order 8 variable density initial under-sampling pattern, and a parallel imaging under-sampling pattern. We next train networks on a 2D polynomial order 8 variable density initial under-sampling pattern at different accelerations factors. Finally, we performed an ablation study by removing the dual domain loss function and the triple reconstruction output to study the effects of the components of the model to performance.

III. RESULTS

Qualitative comparisons between the self-supervised reconstruction methods at R=6 with an initial 2D variable density sampling pattern is shown in Fig. 4. The quantitative metrics for the entire test dataset are shown in the top left of each image. The lower row displays magnified brain regions to better show fine details. All self-supervised methods produce high quality reconstructions with very minimal

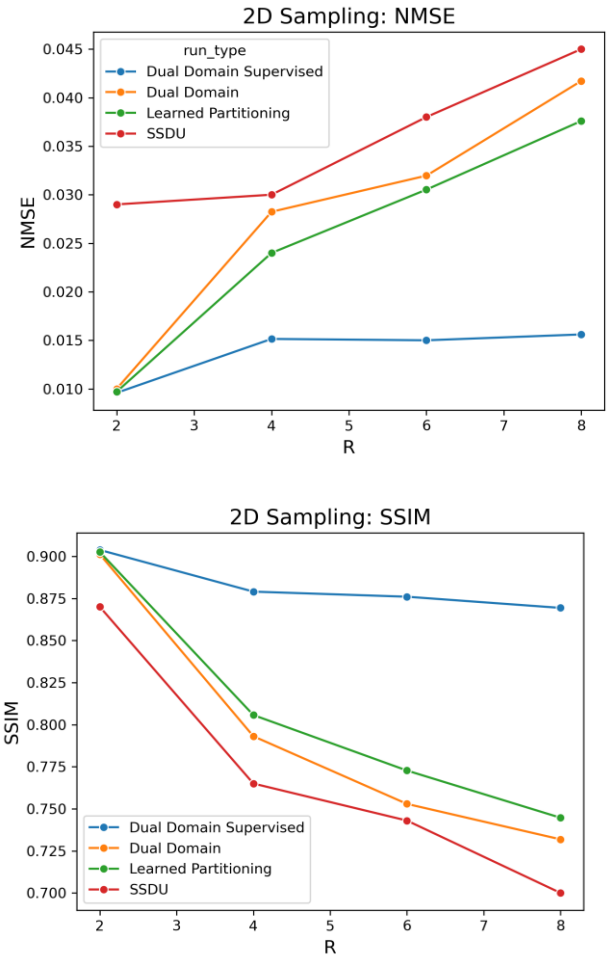

Fig. 5 Quantitative metrics over different acceleration factors. Top: normalized mean squared error metric. Bottom: Structural similarity index.

residual aliasing. Dual domain with learned partitioning resulted in the lowest error and sharper features than the other self-supervised partitioning methods. Quantitative evaluation further supports these findings, with our method achieving the highest SSIM of 0.7273 and the lowest NMSE of 0.030524. Dual domain supervised performs best but has access to fully sampled k-space data.

Quantitative comparisons between self-supervised methods with the remaining initial sampling patterns are shown in Table 1. We compare two different under-sampling distributions: 1D variable density and parallel imaging. Since parallel imaging has a deterministic sampling distribution, Noiser2Noise method [27] used a 2D variable density of R=2 for partitioning. The quantitative metrics for different sampling patterns follows the pervious results with dual domain learned partitioning performing with the lowest error with SSIM values of 0.76959 and 0.76437 for 1D and parallel imaging respectively.

The learned partitioning distribution for each initial sampling distribution at R=6 can be seen in Fig.4. The learned distribution for 2D initial distribution converges to a similar distribution to the M_Ω sampling distribution which is consistent to the theoretical justification for SSDU [27]. The probability distribution differs slightly from M_Ω as the low

TABLE II. QUANTITATIVE ABLATION RESULTS

	SSIM	NMSE
Learned Partitioning SSDU	0.54027	0.12139
$\tilde{y}, \tilde{y}_c$-SSDU	0.74115	0.03832
$\tilde{y}, \tilde{y}_c$-Dual Domain	0.74972	0.03658
Dual Domain Unweighted Loss	0.76170	0.03465
Our Method	**0.77273**	**0.03052**

frequency center region is more rough allowing for low frequency information to pass to both sets. The 1D distribution does not follow the theoretical justification in [27]. For 1D sampling distribution, the learned partitioning resembles a composition between 1D and 2D distributions with the network learning uneven center high frequency regions. For the parallel imaging sampling distribution, there does not exist a theoretically justified partitioning distribution due to deterministic sampling. Since the initial mask is fixed, the network only learns parameters in regions that are sampled. Again, the center of k-space has an uneven texture allowing for low frequency information to path to both sets.

To compare our methods robustness at different acceleration factors, we train self-supervised models at R=2, 4, 6, 8. Quantitative metrics of SSIM and NMSE between self-supervised methods are shown in Fig. 5. Learned partitioning performs with lower error for all acceleration factors and highest SSIM values compared to heuristic partitioning. Supervised has the best performance but has access to fully sampled data.

To show the necessity of model training components, we train 4 ablation models all with learned partitioning. We first trained learned partitioning with the original SSDU loss function (5a) termed learned partitioning SSDU. Next, we add the double k-space loss term (5a+5b) which we name $\tilde{y}$,$\tilde{y}_c$-SSDU. As $f_\theta(y)$ is not needed in the loss term we only reconstruct the input and DC sets. We additionally train a dual domain model with only the input and DC set reconstruction output named $\tilde{y}, \tilde{y}_c$-Dual Domain. For the final ablation, dual domain unweighted loss, we introduce y as a reconstruction output but fix $\alpha = 0.5$ and all β to 7×10^{-7}. Quantitative metrics from all models are shown in Table 2. Applying learned partitioning directly to SSDU results in a drastic performance decrease and yields visibly degraded image quality, reflected in the significantly lower loss. The addition of the reconstruction k-space loss stabilizes training. The addition of the dual domain loss function adds an improvement as both reconstructed images stay consistent. Finally, the introduction of y adds a significant improvement on the quantitative metrics.

IV. DISCUSSION AND CONCLUSION

In this work we developed a learned partitioning network for self-supervised learning which alleviates the need for hand-picked partitioning schemes. This network adds small quantity (~65k) of new parameters compared to SSDU and improves the performance over dual domain self-supervised methods. Our method results in smaller errors for all types of initial sampling distributions and acceleration factors which

show the importance of partitioning in self-supervised learning.

Previous studies have shown that learned sampling improves reconstruction performance by identifying the optimal k-space locations for image reconstruction [11], [12], [13], [33], [34], [35], [36]. Our findings show that learning an optimal partitioning method can also improve reconstruction performance. Near the center of k-space, the learned partitioning distribution is no longer smooth and resembles a Poisson disc sampling pattern with a discrete probability distribution [37]. One possible explanation for converging to a discrete distribution at the center of k-space is that both sets $\tilde{y}$ and $\tilde{y}_c$ get a representative portion of the low frequency data, which carries the majority of the signal intensity. Another possible explanation is that the data around a k-space point is highly correlated due to the receiver coil sensitivities. By learning a Poisson disc pattern near the center, k-space points are evenly spaces so the network can effectively reduce the loss by inferring neighboring k-space points using coil sensitivities. Meanwhile, the regions outside of the low frequency center resemble more of the initial sampling distribution which reflects the findings in the theoretical framework.

The ablation studies show the necessity for the three reconstruction outputs in learning self-supervised partitioning distributions. This requirement can be seen in the learned partitioning SSDU row in Fig 5. When applying LOUPE to learn partitioning in SSDU, the learned partitioning distribution solely prioritizes the center of k-space which contains high energy k-space. This bias is due to the networks ability to trivially reduce the loss function by learning an M_Λ that only calculates the loss on low energy k-space regions. This trivial loss minimization results in a poorly representative partitioning of k-space, ultimately leading to suboptimal reconstruction performance.

The primary limitation of this work is the introduction of new hyperparameters for scaling the k-space and image-space losses. Additionally, dual domain learned partitioning reconstructs three images in each forward pass, resulting in a threefold increase in computational and memory requirements compared to SSDU. Future work will focus on reducing the need for additional hyperparameters, accelerating training, and to further understand the performance benefit in dual domain methods.

REFERENCES

[1] M. A. Griswold *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002, doi: 10.1002/mrm.10171.

[2] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999, doi: 10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S.

[3] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed Sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008, doi: 10.1109/MSP.2007.914728.

[4] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil, "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging," *Magnetic Resonance in Medicine*, vol. 81, no. 1, pp. 439–453, 2019, doi: 10.1002/mrm.27420.

[5] G. Yang *et al.*, "DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018, doi: 10.1109/TMI.2017.2785879.

[6] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," Jul. 16, 2022, *arXiv*: arXiv:2110.05243. Accessed: Jul. 02, 2024. [Online]. Available: http://arxiv.org/abs/2110.05243

[7] G. Luo, M. Blumenthal, M. Heide, and M. Uecker, "Bayesian MRI reconstruction with joint uncertainty estimation using diffusion models," *Magnetic Resonance in Medicine*, vol. 90, no. 1, pp. 295–311, 2023, doi: 10.1002/mrm.29624.

[8] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model Based Deep Learning Architecture for Inverse Problems," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 394–405, Feb. 2019, doi: 10.1109/TMI.2018.2865356.

[9] A. Sriram *et al.*, "End-to-End Variational Networks for Accelerated MRI Reconstruction," Apr. 15, 2020, *arXiv*: arXiv:2004.06688. Accessed: Jan. 09, 2023. [Online]. Available: http://arxiv.org/abs/2004.06688

[10] C. Alkan, M. Mardani, C. Liao, Z. Li, S. S. Vasanawala, and J. M. Pauly, "AutoSamp: Autoencoding k-space Sampling via Variational Information Maximization for 3D MRI," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2024, doi: 10.1109/TMI.2024.3443292.

[11] H. K. Aggarwal and M. Jacob, "J-MoDL: Joint Model-Based Deep Learning for Optimized Sampling and Reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1151–1162, Oct. 2020, doi: 10.1109/JSTSP.2020.3004094.

[12] G. Wang, T. Luo, J.-F. Nielsen, D. C. Noll, and J. A. Fessler, "B-Spline Parameterized Joint Optimization of Reconstruction and K-Space Trajectories (BJORK) for Accelerated 2D MRI," *IEEE Trans. Med. Imaging*, vol. 41, no. 9, pp. 2318–2330, Sep. 2022, doi: 10.1109/TMI.2022.3161875.

[13] C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-Learning-Based Optimization of the Under-Sampling Pattern in MRI," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1139–1152, 2020, doi: 10.1109/TCI.2020.3006727.

[14] H. Sun, C. Tian, J. Xiao, and Y. Yang, "Learn Stable MRI Under-Sampling Pattern With Decoupled Sampling Preference," *IEEE Transactions on Computational Imaging*, vol. 10, pp. 246–260, 2024, doi: 10.1109/TCI.2024.3361773.

[15] J. Zhang *et al.*, "Extending LOUPE for K-Space Under-Sampling Pattern Optimization in Multi-coil MRI," in *Machine Learning for Medical Image Reconstruction*, F. Deeba, P. Johnson, T. Würfl, and J. C. Ye, Eds., Cham: Springer International Publishing, 2020, pp. 91–101. doi: 10.1007/978-3-030-61598-7_9.

[16] Y. Bengio, N. Léonard, and A. Courville, "Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation," Aug. 15, 2013, *arXiv*: arXiv:1308.3432. Accessed: Nov. 04, 2024. [Online]. Available: http://arxiv.org/abs/1308.3432

[17] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," presented at the International Conference on Learning Representations, 2014.

[18] C. Hu, C. Li, H. Wang, Q. Liu, H. Zheng, and S. Wang, "Self-supervised Learning for MRI Reconstruction with a Parallel Network Training Framework," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds., Cham: Springer International Publishing, 2021, pp. 382–391. doi: 10.1007/978-3-030-87231-1_37.

[19] J. Yoo, K. H. Jin, H. Gupta, J. Yerly, M. Stuber, and M. Unser, "Time-Dependent Deep Image Prior for Dynamic MRI," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3337–3348, Dec. 2021, doi: 10.1109/TMI.2021.3084288.

[20] B. Yaman *et al.*, "Multi-mask self-supervised learning for physics-guided neural networks in highly accelerated magnetic resonance imaging," *NMR in Biomedicine*, vol. 35, no. 12, p. e4798, 2022, doi: 10.1002/nbm.4798.

[21] M. Blumenthal, C. Fantinato, C. Unterberg-Buchwald, M. Haltmeier, X. Wang, and M. Uecker, “Self-supervised learning for improved calibrationless radial MRI with NLINV-Net,” *Magnetic Resonance in Medicine*, vol. 92, no. 6, pp. 2447–2463, 2024, doi: 10.1002/mrm.30234.

[22] P. Hua *et al.*, “Deep MRI Reconstruction without Ground Truth for Training,” in *ISMRM annual meeting*, May 2019.

[23] Y. Korkmaz, T. Cukur, and V. M. Patel, “Self-supervised MRI Reconstruction with Unrolled Diffusion Models,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, H. Greenspan, A. Madabhushi, P. Mousavi, S. Salcudean, J. Duncan, T. Syeda-Mahmood, and R. Taylor, Eds., in Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023, pp. 491–501. doi: 10.1007/978-3-031-43999-5_47.

[24] P. Huang, C. Zhang, X. Zhang, X. Li, L. Dong, and L. Ying, “Self-Supervised Deep Unrolled Reconstruction Using Regularization by Denoising,” *IEEE Transactions on Medical Imaging*, vol. 43, no. 3, pp. 1203–1213, Mar. 2024, doi: 10.1109/TMI.2023.3332614.

[25] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3172–3191, 2020, doi: 10.1002/mrm.28378.

[26] B. Zhou *et al.*, “Dual-domain self-supervised learning for accelerated non-Cartesian MRI reconstruction,” *Medical Image Analysis*, vol. 81, p. 102538, Oct. 2022, doi: 10.1016/j.media.2022.102538.

[27] C. Millard and M. Chiew, “A Theoretical Framework for Self-Supervised MR Image Reconstruction Using Sub-Sampling via Variable Density Noisier2Noise,” *IEEE Transactions on Computational Imaging*, vol. 9, pp. 707–720, 2023, doi: 10.1109/TCI.2023.3299212.

[28] N. Moran, D. Schmidt, Y. Zhong, and P. Coady, “Noisier2Noise: Learning to Denoise From Unpaired Noisy Data,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 12061–12069. doi: 10.1109/CVPR42600.2020.01208.

[29] M. Blumenthal, G. Luo, M. Schilling, M. Haltmeier, and M. Uecker, “Nlinv-net: Self-supervised end-2-end learning for reconstructing undersampled radial cardiac real-time data,” in *Proceedings of the International Society for Magnetic Resonance in Medicine*, May 2022.

[30] C. Liu, P. Hu, and H. Qi, “Deep-learning-based optimization of k-space undersampling in self-supervised MRI reconstruction,” in *ISMRM annual meeting*, May 2024.

[31] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Jan. 30, 2017, *arXiv*: arXiv:1412.6980. Accessed: Nov. 05, 2024. [Online]. Available: http://arxiv.org/abs/1412.6980

[32] J. Zbontar *et al.*, “fastMRI: An Open Dataset and Benchmarks for Accelerated MRI,” Dec. 11, 2019, *arXiv*: arXiv:1811.08839. Accessed: Feb. 28, 2023. [Online]. Available: http://arxiv.org/abs/1811.08839

[33] J. Yang *et al.*, “Fast Multi-Contrast MRI Acquisition by Optimal Sampling of Information Complementary to Pre-Acquired MRI Contrast,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1363–1373, May 2023, doi: 10.1109/TMI.2022.3227262.

[34] B. Gözcü *et al.*, “Learning-Based Compressive MRI,” May 03, 2018, *arXiv*: arXiv:1805.01266. Accessed: Jan. 15, 2024. [Online]. Available: http://arxiv.org/abs/1805.01266

[35] M. Seeger, H. Nickisch, R. Pohmann, and B. Schölkopf, “Optimization of k-space trajectories for compressed sensing by Bayesian experimental design,” *Magnetic Resonance in Medicine*, vol. 63, no. 1, pp. 116–126, 2010, doi: 10.1002/mrm.22180.

[36] C. Lazarus *et al.*, “SPARKLING: variable-density k-space filling curves for accelerated T2*-weighted MRI,” *Magnetic Resonance in Medicine*, vol. 81, no. 6, pp. 3643–3661, 2019, doi: 10.1002/mrm.27678.

[37] M. Lustig and J. M. Pauly, “SPIRiT: Iterative Self-consistent Parallel Imaging Reconstruction from Arbitrary k-Space,” *Magn Reson Med*, vol. 64, no. 2, pp. 457–471, Aug. 2010, doi: 10.1002/mrm.22428.