Efficient Noise Calculation in Deep Learning-based MRI Reconstructions

Onat Dalmaz [1 2] **Arjun D. Desai** [1 2] **Reinhard Heckel** [3] **Tolga Cukur** [4] **Akshay S. Chaudhari** [2 5]
Brian Hargreaves [1 2 6]

Abstract

Accelerated MRI reconstruction involves solving an ill-posed inverse problem where noise in acquired data propagates to the reconstructed images. Noise analyses are central to MRI reconstruction for providing an explicit measure of solution fidelity and for guiding the design and deployment of novel reconstruction methods. However, deep learning (DL)-based reconstruction methods have often overlooked noise propagation due to inherent analytical and computational challenges, despite its critical importance. This work proposes a theoretically grounded, memory-efficient technique to calculate *voxel-wise variance* for quantifying uncertainty due to acquisition noise in accelerated MRI reconstructions. Our approach approximates noise covariance using the DL network's Jacobian, which is intractable to calculate. To circumvent this, we derive an *unbiased estimator* for the diagonal of this covariance matrix—voxel-wise variance—, and introduce a Jacobian sketching technique to efficiently implement it. We evaluate our method on knee and brain MRI datasets for both data- and physics-driven networks trained in supervised and unsupervised manners. Compared to empirical references obtained via Monte-Carlo simulations, our technique achieves near-equivalent performance while reducing computational and memory demands by an order of magnitude or more. Furthermore, our method is robust across varying input noise levels, acceleration factors, and diverse undersampling schemes, highlighting its broad applicability. Our work *reintroduces* accurate and efficient noise analysis as a central tenet of reconstruction algorithms, holding promise to reshape how we evaluate and deploy DL-based MRI.

1. Introduction

Magnetic Resonance Imaging (MRI) has become indispensable in clinical diagnostics, yet prolonged acquisition times can reduce patient throughput and exacerbate motion artifacts. Parallel imaging (pMRI) (Pruessmann et al., 1999; Griswold et al., 2002) was developed to mitigate these challenges by undersampling *k*-space and exploiting multiple receiver coils with distinct sensitivity profiles. Despite successfully reducing scan duration, pMRI inevitably introduces spatially varying noise amplification—commonly quantified by the *g-factor*—due to the need to invert an ill-conditioned system. In regions where coil sensitivities overlap or are highly attenuated, the reconstruction leads to locally elevated noise levels (Pruessmann et al., 1999; Sodickson et al., 1999). Early pMRI research rigorously analyzed this spatial noise amplification, culminating in precise formulations that ensured reliable signal-to-noise ratio (SNR) performance to guide algorithm design and clinical implementation (Pruessmann et al., 1999; Sodickson et al., 1999; Aja-Fernández et al., 2014). For instance, the widely used SENSE algorithm explicitly incorporates coil sensitivities to minimize noise, reflecting how classical pMRI approaches typically maintain a clear link between measurement noise and the reconstructed image.

In contrast, most deep learning (DL) reconstruction methods (Akçakaya et al., 2019; Knoll et al., 2020a; Yaman et al., 2022; Jalal et al., 2021; Jun et al., 2021; Wang et al., 2024; Sriram et al., 2020b) have not provided an explicit account of how noise from undersampled k-space propagates into the final image. A major obstacle lies in the highly *nonlinear* and complex nature of DL reconstructions (Chen et al., 2022) making it nontrivial to derive analytical frameworks to characterize how undersampling and learned regularization affect localized noise amplification. In response, the core emphasis in the design of DL methods typically sought maximization of summary metrics like peak signal-to-noise ratio (PSNR) or structural similarity index (SSIM) that do

[1]Department of Electrical Engineering, Stanford University, Stanford, CA, USA [2]Department of Radiology, Stanford University, Stanford, CA, USA [3]Department of Computer Engineering, Technical University of Munich, Munich, Germany [4]Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey [5]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA [6]Department of Bioengineering, Stanford University, Stanford, CA, USA. Correspondence to: Onat Dalmaz <onat@stanford.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

not provide an acute assessment on the spatial noise distribution (Heckel et al., 2024; Adamson et al., 2023; Muckley et al., 2021). Thus, the interplay between noise propagation and neural network parameters—and its influence on reconstruction robustness—has received comparatively little attention (Darestani et al., 2021; Dalmaz et al., 2024).

Nevertheless, characterizing noise in reconstructions is pivotal for advancing DL-MRI, as noise directly impacts SNR—a fundamental image quality metric in MRI that strongly influences clinical utility (Breuer et al., 2009; Dietrich et al., 2007). Although metrics such as PSNR and SSIM remain common in DL, these metrics provide *image-wide summary measures* of performance and do not capture local variations in noise or SNR across the image (Mason et al., 2020). Consequently, important diagnostic regions may be obscured by unpredictable noise amplification patterns that that are either invisible or averaged out when using these global metrics (Adamson et al., 2021; Knoll et al., 2020b). In contrast, explicit noise quantification by means of spatial variance maps can yield critical insights into how reconstruction algorithms handle and potentially amplify noise, shaping design decisions to improve robustness and reliability (Kellman & McVeigh, 2005; Knoll et al., 2019), paving the way for clinically meaningful performance improvements.

In modern unsupervised (Yaman et al., 2020; Xiang et al., 2023) or semi-supervised paradigms (Yurt et al., 2022), SNR metrics can be especially valuable by enabling benchmarking of reconstructions without relying on fully sampled ground-truth data (Kastryulin et al., 2023). Moreover, understanding noise propagation can inform tailored sampling schemes—enabling the identification of k-space points that contribute disproportionately to image variability (Alkan et al., 2024; Peng et al., 2022) and guide development of noise-aware architectures and training strategies (Desai et al., 2023). Additionally, accurate noise characterization benefits downstream imaging tasks—such as denoising, segmentation, and diagnostic analysis—by providing voxel-wise variance estimates that enhance algorithm robustness and clinical interpretability (Dou et al., 2025; Wei et al., 2022). Ultimately, noise characterization strengthens image-quality assessments, fosters reliable DL-based MRI algorithms, and increases clinical confidence in deployed models (Chaudhari et al., 2021).

Our contributions are threefold: (i) We provide a comprehensive theoretical framework that provides insights into how acquisition noise in k-space propagates to image uncertainty in DL-based MRI reconstructions. *Voxel-wise variance* of reconstructed images is linked to row vectors of DL network's Jacobian, modulated by k-space correlations and the imaging operator. To circumvent calculation of the full Jacobian, an unbiased estimator for the diagonal elements of the noise covariance matrix is introduced; **(ii)** We show how the derived estimator can be efficiently implemented through a novel Jacobian sketching algorithm that leverages a complex-valued sketching matrix with random-phase columns, modulated by the adjoint operator and the noise covariance. This approach offers significant advantages over traditional Monte-Carlo (MC) simulation-based methods by reducing computational and memory demands by an order of magnitude or more; **(iii)** We rigorously evaluate the proposed method on knee and brain MRI datasets using various network architectures, including data-driven and physics-driven models trained in paradigms ranging from supervised to unsupervised. Experimental results demonstrate that the technique matches MC simulation-based empirical references in noise calculation with average correlation 99.8% and error 0.8%. We further validate effectiveness of our method across different input noise levels, acceleration factors, and undersampling schemes, showing robustness across clinically relevant imaging scenarios. Collectively, our method aims to reestablish accurate and efficient noise analysis in reconstruction algorithms, enhancing the evaluation and deployment of DL-based MRI.

2. Related Work

Jacobian Sketching in Machine Learning. Jacobian sketching has been utilized in machine learning for diverse objectives, such as mitigating catastrophic forgetting in continual learning (Heckel, 2022; Li et al., 2021) or reducing the variance of stochastic quasi-gradient methods (Gower et al., 2021). In such applications, the primary aim is typically to approximate Jacobian- or Hessian-related quantities to improve model optimization or regularization. Our work, however, leverages Jacobian sketching for a fundamentally different purpose: computationally efficient *quantitative analysis of noise propagation* in nonlinear multi-coil deep MRI reconstructions. Specifically, we employ sketching to efficiently estimate the diagonal elements of the image covariance matrix (i.e., voxel-wise noise variances) arising from propagation of the measurement noise, rather than optimization or regularization objectives.

Noise Calculation in DL-MRI. Related work on quantifying noise in DL-based MRI reconstructions often employs MC simulations (Robson et al., 2008; Thunberg & Zetterberg, 2007; Akcakaya et al., 2014; Dalmaz et al., 2024), repeatedly injecting synthetic perturbations into k-space and reconstructing numerous noisy realizations. This approach can produce accurate estimates via a large number of trials which require considerable computational and memory resources—and is thus impractical for large 3D or 4D volumes. Moreover, treating the network as a black box limits interpretability by obscuring how noise propagates or is amplified within the deep reconstruction pipeline. A

recent theoretical result (Dawood et al., 2024) derives analytical noise estimates for *k*-space interpolation networks, and thereby is not generalizable across different architectures. Furthermore, it involves computing the entire network Jacobian, which quickly becomes impractical for typical 2D or 3D MRI volumes. These limitations highlight the need for a versatile, scalable, and theoretically grounded approach to noise analysis—one that avoids intensive sampling, preserves interpretability, and can be seamlessly integrated into diverse reconstruction paradigms.

Uncertainty Quantification in MRI. Prior works on uncertainty quantification (UQ) in MRI generally fall into three categories: (1) classical compressed sensing approaches, often using Monte Carlo (MC) reconstructions from perturbed measurements to derive confidence intervals (Hoppe et al., 2024); (2) Bayesian methods (Narnhofer et al., 2021) or generative modeling techniques (Tezcan et al., 2020), such as Variational Autoencoders (VAEs) or those assuming i.i.d. Gaussian priors, to capture uncertainty, primarily epistemic; and (3) ensemble-based analyses focusing on epistemic uncertainty, for instance, by training multiple networks in parallel (Küstner et al., 2024).

Some recent approaches utilize a conformal prediction framework to construct distribution-free uncertainty intervals at the level of downstream task outputs (Wen et al., 2024). However, this method treats the reconstruction pipeline as a black box, does not model voxel-level uncertainty intrinsically, does not incorporate the physics of MRI acquisition, and requires calibration data to guarantee finite-sample statistical coverage; it does not estimate how acquisition noise propagates through multi-coil systems or nonlinear reconstructions. Another distinct approach by Edupuganti et al. (Edupuganti et al., 2021) introduced a VAE-based probabilistic framework that models epistemic reconstruction uncertainty in a latent space. It utilizes SURE-based estimators—relying on approximate Jacobian traces—and MC sampling to compute uncertainty maps, but under the assumption of uncorrelated, i.i.d., single-coil noise.

Our work uniquely closes a longstanding gap by providing the first rigorous, first-principles derivation of how multi-coil *k*-space acquisition noise propagates through both MRI physics and nonlinear deep learning (DL) networks. While exact voxel-wise variance requires the full, computationally intractable network Jacobian, our core innovation is a statistically rigorous, unbiased estimator for this variance. This estimator, derived from imaging and statistical theory, uniquely handles complex-valued signals, multi-coil encoding, physics-based models, and DL nonlinearities simultaneously. We implement this via a *practical, model-agnostic matrix sketching algorithm using Jacobian-vector products (JVPs) with random-phase vectors*, achieving scalable, interpretable, and memory-efficient voxel-wise noise variance (NV) estimates. This targets the *aleatoric* uncertainty due to acquisition noise, offering a mathematically transparent solution distinct from prior UQ and noise calculation methods.

3. Theory

3.1. Accelerated MRI

We acquire *k*-space measurements $\boldsymbol{y} \in \mathbb{C}^m$ from an unknown image $\boldsymbol{x} \in \mathbb{C}^n$ via a linear *imaging operator* $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, comprising coil sensitivity maps, Fourier encoding, and sampling mask. In practice, $\boldsymbol{y}$ is often undersampled, and corrupted by noise $\boldsymbol{n} \in \mathbb{C}^m$. Thus, the forward model is:

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{n}. \tag{1}$$

Acquisition noise $\boldsymbol{n}$ is modeled as complex *Additive White Gaussian Noise (AWGN)* with zero mean:

$$\boldsymbol{n} \sim \mathcal{CN}\big(\mathbf{0}, \boldsymbol{\Sigma}_k\big), \tag{2}$$

where $\boldsymbol{\Sigma}_k \in \mathbb{C}^{m \times m}$ is the *sample covariance matrix* of the *k*-space data. At each *k*-space point, noise is correlated across coils but is independent across distinct *k*-space locations. This implies that a *coil covariance matrix* $\widetilde{\boldsymbol{\Sigma}}_k$ fully captures the noise statistics. Yet, for simplicity in our linear-algebraic treatment and factorizations, we focus on the Hermitian positive semi-definite (HPSD) *sample covariance matrix* $\boldsymbol{\Sigma}_k \in \mathbb{C}^{m \times m}$ (see Appendix A).

3.2. Neural Network MRI Reconstruction

In practice, deep MRI reconstructions often follow one of two paradigms: **(1) Unrolled Architectures**, where a series of learned regularization blocks and data-consistency (DC) steps are iterated for K cycles (Hammernik et al., 2023; Liang et al., 2020; Fabian et al., 2022), or **(2) Purely Data-Driven Mappings**, where a feed-forward network performs a direct mapping (Mardani et al., 2019)). Despite their structural differences, both networks can be viewed as a nonlinear function $f(\boldsymbol{A}^H\,\boldsymbol{y})$, where $\boldsymbol{A}^H$ is the adjoint operator that yields a naive zero-filled (ZF) estimate. Without loss of generality, any least-squares or pseudo-inverse initialization could be used (e.g. $\boldsymbol{A}^\dagger\,\boldsymbol{y}$); such choices would simply alter the linear operator involved in subsequent derivations. Thus, our analyses are agnostic to these architectural details, and simply require that f be differentiable so that its Jacobian exists (See Appendix B).

3.3. Noise Propagation

In MRI, we seek to quantify the *voxel-wise variance* of the reconstructed image $\boldsymbol{x} \in \mathbb{C}^n$ induced by the acquisition

noise in k-space. Let $f : \mathbb{C}^n \to \mathbb{C}^n$ denote a learned *reconstruction function* (e.g., an unrolled network), with Jacobian $\boldsymbol{J}_f(\boldsymbol{x}) = \left[\begin{array}{ccc} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{array} \right] \in \mathbb{C}^{n \times n}$.

DL pipelines commonly initialize $\boldsymbol{x}^{(0)} = \boldsymbol{A}^H \boldsymbol{y}$ with the "noisy" measured data $\boldsymbol{y} = \boldsymbol{y}_0 + \boldsymbol{n}$, where $\boldsymbol{y}_0 = \boldsymbol{A}\,\boldsymbol{x}$. Then:

$$\boldsymbol{x}^{(0)} = \boldsymbol{A}^H \left(\boldsymbol{y}_0 + \boldsymbol{n}\right) = \boldsymbol{A}^H \boldsymbol{y}_0 + \boldsymbol{A}^H \boldsymbol{n} = \boldsymbol{x}_0 + \boldsymbol{A}^H \boldsymbol{n}, \tag{3}$$

with $\boldsymbol{x}_0 = \boldsymbol{A}^H \boldsymbol{y}_0$ indicating the noise-free component of this initial estimate. A *first-order Taylor expansion* of the reconstruction function $f(\cdot)\colon \mathbb{C}^n \to \mathbb{C}^n$ around $\boldsymbol{x}_0$ approximates:

$$f\big(\boldsymbol{x}^{(0)}\big) \;\approx\; f(\boldsymbol{x}_0) \;+\; \boldsymbol{J}_f(\boldsymbol{x}_0)\,\big(\boldsymbol{x}^{(0)} - \boldsymbol{x}_0\big), \tag{4}$$

where $\boldsymbol{J}_f(\boldsymbol{x}_0) \in \mathbb{C}^{n \times n}$ is the Jacobian matrix of f evaluated at $\boldsymbol{x}_0$. From 3 and 4 the perturbation in the reconstructed image due to the noise $\boldsymbol{n}$ is:

$$\delta\boldsymbol{x} = f\big(\boldsymbol{x}^{(0)}\big) - f(\boldsymbol{x}_0) \;\approx\; \boldsymbol{J}_f(\boldsymbol{x}_0)\,\big(\boldsymbol{A}^H \boldsymbol{n}\big)$$

Note that $\delta\boldsymbol{x}$ also remains zero-mean:

$$\mathbb{E}\big[\delta\boldsymbol{x}\big] \;=\; \boldsymbol{J}_f(\boldsymbol{x}_0)\,\boldsymbol{A}^H \mathbb{E}\big[\boldsymbol{n}\big] \;= \mathbf{0}. \tag{5}$$

It is also worthwhile to note that due to linearity, approximate reconstruction noise shares the same Gaussian nature, with the covariance matrix:

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{x}} &= \mathbb{E}\big[\delta\boldsymbol{x}\,\delta\boldsymbol{x}^H\big] = \boldsymbol{J}_f(\boldsymbol{x}_0)\,\boldsymbol{A}^H\,\mathbb{E}[\boldsymbol{n}\,\boldsymbol{n}^H]\,\boldsymbol{A}\,\boldsymbol{J}_f(\boldsymbol{x}_0)^H \\ &= \boldsymbol{J}_f(\boldsymbol{x}_0)\,\boldsymbol{A}^H\,\boldsymbol{\Sigma}_k\,\boldsymbol{A}\,\boldsymbol{J}_f(\boldsymbol{x}_0)^H. \end{aligned} \tag{6}$$

Even though the Jacobian $\boldsymbol{J}_f = \boldsymbol{J}_f(\boldsymbol{x}_0) \in \mathbb{C}^{n \times n}$ exists, storing or explicitly computing the full matrix is intractable for MR images, due to dimensionality and massive data size. Nonetheless, (6) shows propagation of k-space covariance $\boldsymbol{\Sigma}_k$ through the adjoint operator $\boldsymbol{A}^H$ and the network's Jacobian $\boldsymbol{J}_f$. Similar constructs for the noise matrix exist in other pMRI approaches (Pruessmann et al., 1999).

Factorization and Diagonal Entries. To simplify computations, we first perform Cholesky decomposition of $\boldsymbol{\Sigma}_k$:

$$\boldsymbol{\Sigma}_k = \boldsymbol{\sigma}_k\,\boldsymbol{\sigma}_k^H, \tag{7}$$

where $\boldsymbol{\sigma}_k$ is the Cholesky factor of $\boldsymbol{\Sigma}_k$. Substituting into (6), we obtain

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{J}_f \boldsymbol{A}^H \big(\boldsymbol{\sigma}_k \boldsymbol{\sigma}_k^H\big) \boldsymbol{A} \boldsymbol{J}_f^H, \tag{8}$$

$$= \Big(\boldsymbol{J}_f\,\boldsymbol{A}^H\,\boldsymbol{\sigma}_k\Big)\Big(\boldsymbol{J}_f\,\boldsymbol{A}^H\,\boldsymbol{\sigma}_k\Big)^H. \tag{9}$$

Here, we see that $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{L}\,\boldsymbol{L}^H$, with Cholesky factor:

$$\boldsymbol{L} \;=\; \boldsymbol{J}_f\,\boldsymbol{A}^H\,\boldsymbol{\sigma}_k, \tag{10}$$

The variance of the i-th voxel (*i.e.*, the i-th diagonal entry of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$) is then:

$$\mathrm{Var}(\boldsymbol{x}_i) \;=\; \mathrm{diag}(\boldsymbol{\Sigma}_{\boldsymbol{x}})_i \;=\; [\boldsymbol{L}\,\boldsymbol{L}^H]_{ii} \;=\; \|\boldsymbol{l}_i\|_2^2, \tag{11}$$

where $\boldsymbol{l}_i \in \mathbb{C}^n$ is the i-th *row* of $\boldsymbol{L}$:

$$\boldsymbol{l}_i = \nabla^\top f_i\,\boldsymbol{A}^H\,\boldsymbol{\sigma}_k$$

In principle, the variance of each voxel could be iteratively obtained via taking the ℓ_2 norm of i-th row of Jacobian, transformed and scaled by $\boldsymbol{A}^H$ and $\boldsymbol{\sigma}_k$. Although this direct approach would be computationally heavy (see **Algorithm** 1 in Appendix), the analysis reveals how $\boldsymbol{\Sigma}_k$, $\boldsymbol{A}$ and $\boldsymbol{J}_f$ together determine the voxel-wise noise distribution.

3.4. Estimating Noise via Jacobian Sketching

Here we present an algorithm for estimating the diagonal of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ in an unrolled MRI reconstruction, *without* explicitly forming or storing the Jacobian $\boldsymbol{J}_f \in \mathbb{C}^{n \times n}$ or the image covariance $\boldsymbol{\Sigma}_{\boldsymbol{x}} \in \mathbb{C}^{n \times n}$. We instead rely on *Jacobian-Vector Products (JVPs)*, enabling a *sketch-based* approach that efficiently probes the Jacobian and covariance structure through random vectors. We first present the following results regarding the diagonal entries of implicit, complex-valued Hermitian covariance matrices.

Theorem 3.1 (UNBIASED DIAGONAL ESTIMATOR)**.** *Let $\boldsymbol{\Sigma} \in \mathbb{C}^{n \times n}$ be Hermitian, and let $\boldsymbol{v} \in \mathbb{C}^n$ satisfy*

$$\mathbb{E}[\boldsymbol{v}] \;=\; \mathbf{0}, \quad \mathbb{E}[\boldsymbol{v}\,\boldsymbol{v}^H] \;=\; \boldsymbol{I}. \tag{12}$$

Define $\boldsymbol{y} = (\boldsymbol{\Sigma}\,\boldsymbol{v}) \odot \boldsymbol{v}^*$, where $\odot$ is the Hadamard product, and $*$ is scalar complex conjugation. Then:

$$\mathbb{E}[\boldsymbol{y}] = \mathrm{diag}\big(\boldsymbol{\Sigma}\big)$$

Proof. Write $\boldsymbol{\Sigma} = [\Sigma_{ij}]$ and $\boldsymbol{v} = (v_1, \ldots, v_n)^\top$. Then for each index i:

$$(\boldsymbol{\Sigma}\,\boldsymbol{v})_i \;=\; \sum_{j=1}^{n} \Sigma_{ij}\,v_j.$$

Multiplying by v_i^* yields y_i:

$$y_i = (\boldsymbol{\Sigma}\,\boldsymbol{v})_i\,v_i^* \;=\; \sum_{j=1}^{n} \Sigma_{ij}\,v_j\,v_i^*.$$

Taking expectation:

$$\mathbb{E}[y_i] \;=\; \mathbb{E}\Big[\sum_{j=1}^{n} \Sigma_{ij}\,v_j\,v_i\big)^*\Big] \;=\; \sum_{j=1}^{n} \Sigma_{ij}\,\mathbb{E}\Big[v_j\,v_i^*\Big]$$

From 12, we note $\mathbb{E}[\,v_j\,v_i^*\,] = \delta_{ij}$:

$$\mathbb{E}[y_i] = \sum_{j=1}^{n} \Sigma_{ij}\,\delta_{ij} \;=\; \Sigma_{ii}.$$

Since this holds for all i, we have $\mathbb{E}[\boldsymbol{y}] = \mathrm{diag}(\boldsymbol{\Sigma})$. □

Lemma 3.2. *Let $\boldsymbol{\Sigma} \in \mathbb{C}^{n\times n}$ be HPSD with a Cholesky factorization $\boldsymbol{\Sigma} = \boldsymbol{L}\,\boldsymbol{L}^H$. Suppose $\boldsymbol{v} \in \mathbb{C}^n$ satisfies (12). Defining $\boldsymbol{u} = \boldsymbol{L}\,\boldsymbol{v}$, we claim:*

$$\mathbb{E}\big[\boldsymbol{u} \odot \boldsymbol{u}^*\big] = \mathrm{diag}\big(\boldsymbol{\Sigma}\big).$$

Proof. Since $\boldsymbol{u} = \boldsymbol{L}\,\boldsymbol{v}$, its ith component is

$$u_i = \sum_{j=1}^{n} L_{ij}\, v_j.$$

Hence, the ith entry of the Hadamard product $\boldsymbol{u} \odot \boldsymbol{u}^*$ is

$$[\boldsymbol{u} \odot \boldsymbol{u}^*]_i \;=\; u_i\,(u_i)^* \;=\; \Big(\sum_{j=1}^{n} L_{ij}\, v_j\Big)\Big(\sum_{k=1}^{n} L_{ik}^*\, v_k^*\Big).$$

Rewriting and taking expectation:

$$\mathbb{E}\big[u_i\,(u_i)^*\big] = \sum_{j=1}^{n}\sum_{k=1}^{n} L_{ij}\, L_{ik}^*\, \mathbb{E}[v_j\, v_k^*].$$

By (12), $\mathbb{E}[v_j\, v_k^*] = \delta_{jk}$. Thus,

$$\mathbb{E}\big[u_i\,(u_i)^*\big] = \sum_{j=1}^{n} L_{ij}\, L_{ij}^* = [\boldsymbol{L}\,\boldsymbol{L}^H]_{ii} = \Sigma_{ii}.$$

Because this holds for each i, we conclude

$$\mathbb{E}[\boldsymbol{u} \odot \boldsymbol{u}^*] = \mathrm{diag}(\boldsymbol{\Sigma}).$$

□

Returning to (10), we have, $\boldsymbol{L} = \boldsymbol{J}_f\,\boldsymbol{A}^H\,\boldsymbol{\sigma}_k$. Then, by Lemma 3.2,

$$\boldsymbol{u} = \boldsymbol{L}\,\boldsymbol{v} = \boldsymbol{J}_f\Big(\boldsymbol{A}^H(\boldsymbol{\sigma}_k\,\boldsymbol{v})\Big) \Rightarrow \mathbb{E}\big[\boldsymbol{u} \odot \boldsymbol{u}^*\big] = \mathrm{diag}(\boldsymbol{\Sigma}_{\boldsymbol{x}}).$$

Thus, applying $\boldsymbol{J}_f$ to suitably distributed random vectors reveals each diagonal entry of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$.

3.4.1. Vectorized Implementation: Jacobian Sketching

A naive approach would sample each $\boldsymbol{v}_j \in \mathbb{C}^m$, compute $\boldsymbol{L}\,\boldsymbol{v}_j = \boldsymbol{J}_f(\boldsymbol{A}^H\boldsymbol{\sigma}_k\,\boldsymbol{v}_j)$ individually, and accumulate $\boldsymbol{u}_j$. Instead, we provide a *vectorized* algorithm that improves computational and efficiency (See Algorithm 2 in Appendix):

1. **Generate random matrix:** $\boldsymbol{V}_S \in \mathbb{C}^{m\times S}$ with columns $\boldsymbol{v}_j$ s.t. $\mathbb{E}[\boldsymbol{v}_j] = \boldsymbol{0}$, and $\mathbb{E}[\boldsymbol{v}_j\,\boldsymbol{v}_j^H] = \boldsymbol{I}_m$.
2. **Transform by $\boldsymbol{\sigma}_k$ and $\boldsymbol{A}^H$:** $\boldsymbol{W}_S = \boldsymbol{\sigma}_k\,\boldsymbol{V}_S$, then $\widetilde{\boldsymbol{W}}_S = \boldsymbol{A}^H\,\boldsymbol{W}_S$
3. **Sketch $\boldsymbol{J}_f$ via $\widetilde{\boldsymbol{W}}_S$:** $\boldsymbol{U}_S = \boldsymbol{J}_f\widetilde{\boldsymbol{W}}_S \in \mathbb{C}^{n\times S}$.
4. **Hadamard product w/Hermitian & Average:** $\boldsymbol{V}_{\text{samples}} = \boldsymbol{U}_S \odot \boldsymbol{U}_S^H \in \mathbb{R}^{n\times S}$, $\widehat{\mathrm{diag}(\boldsymbol{\Sigma}_{\boldsymbol{x}})} = \frac{1}{S}\,\boldsymbol{V}_{\text{samples}}\,\boldsymbol{1}_S$.

3.4.2. Choice of Random Vectors

For unbiased estimation, we require random vectors $\boldsymbol{v} \in \mathbb{C}^n$ satisfying (12). A natural design choice would be *standard complex Gaussian* vectors where each v_i is drawn from $\sim \mathcal{CN}(0,1)$ independently. Yet, in *real*-valued matrix diagonal estimation problems, *Rademacher* vectors (i.e., ± 1) have been shown to reduce estimator variance compared to Gaussian vectors (Hutchinson, 1990; Bekas et al., 2007). In MR reconstruction, however, data and operators are inherently *complex*, thus we propose using a *complex* analogue, namely *complex Rademacher (random-phase)* vectors, where each v_i has unit magnitude and uniformly random phase:

$$v_i = e^{j\,\theta_i},\ \theta_i \sim \mathrm{Uniform}[0, 2\pi].$$

We show in Appendix D that both choices satisfy (12), though *random-phase* vectors yield strictly lower *estimator variance*; ergo, selected for consequent experiments. Theoretical error bounds for the estimator using random-phase vectors are provided in Appendix E.

4. Experiments and Results

4.1. Datasets and Experimental Setup

We performed experiments on two publicly available datasets—**Stanford knee dataset** and a subset of the **fastMRI brain dataset** (Knoll et al., 2020c). The Stanford knee dataset consists of 8-channel 3D FSE PD-weighted scans, which we split into 14 subjects for training, 2 for validation, and 3 for testing. Each 3D scan was demodulated and decoded via a 1D inverse Fourier transform along the readout dimension, yielding 2D Axial slices with matrix size 320×256. The fastMRI brain dataset comprises 16-channel Axial T2-weighted scans with matrix size 384×384, split into 54 training, 20 validation, and 30 testing subjects. We focus on using these test volumes primarily to evaluate the variance estimation performance of our framework, rather than to compare the reconstruction quality of different methods. k-space data were retrospectively undersampled via 2D Poisson Disc undersampling masks; and coil sensitivities were estimated via J-SENSE (Ying & Sheng, 2007). We estimated the coil covariance matrix $\widetilde{\boldsymbol{\Sigma}}_k$ for each slice using the outermost 5% of k-space points, where signals are minimal (Robson et al., 2008; Pruessmann et al., 1999). During inference, a unique and deterministic undersampling mask was used on each test slice for reproducibility.

Reconstruction Methods To comprehensively evaluate our approach, we selected six deep MRI reconstruction algorithms spanning different learning paradigms: supervised (**E2E-VarNet** (Sriram et al., 2020a), **MoDL** (Aggarwal et al., 2019), **U-Net** (Ronneberger et al., 2015)), semi-supervised (**N2R (Noise2Recon)**, **VORTEX** (Desai

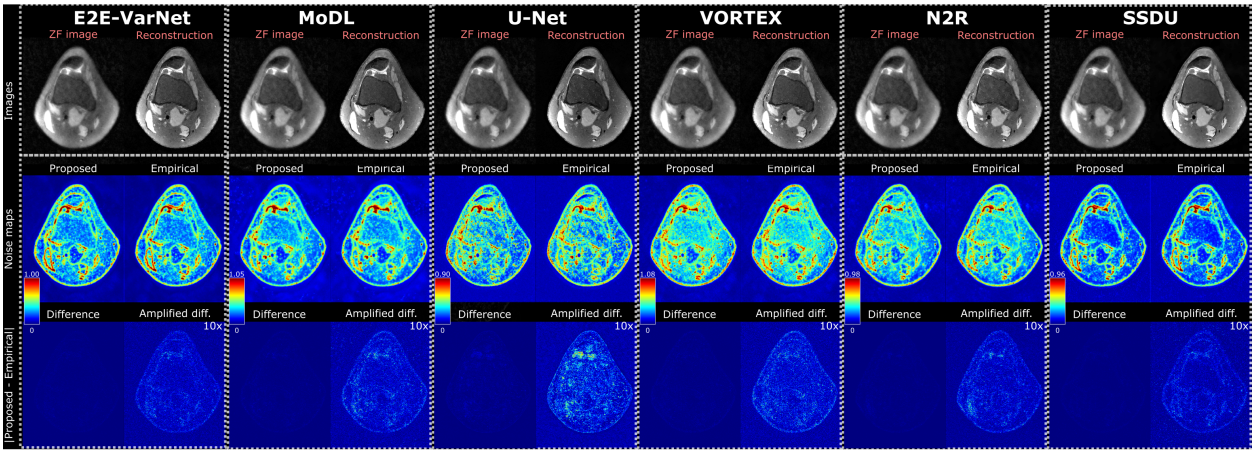

Figure 1. Each column corresponds to a distinct deep reconstruction method at $R = 8, \alpha = 1$ on knee data. In each column: **(top row)** shows ZF and reconstructed images; **(middle row)** displays NV maps derived by the proposed method and empirical simulations; **(bottom row)** presents difference and amplified (x10) difference maps between the proposed and empirical variance maps to highlight spatial discrepancies. Color bars indicate each panel's relative noise map display window.

et al., 2022a)), self-supervised (**SSDU** (Yaman et al., 2020)); and network architectures: fully data-driven (**U-Net** (Ronneberger et al., 2015)) vs. physics-driven unrolled (the rest). These methods cover a broad spectrum of deep reconstruction models for examining the generalizability of the proposed noise analysis framework across different architectures and learning paradigms. Additional details regarding methods including training, hyperparameters, architectures, and results are provided in Appendix H.

Simulating Noise Levels. To investigate performance under varying noise scenarios, we simulate different input noise levels by scaling the estimated covariance matrix $\boldsymbol{\Sigma}_k = \alpha\widetilde{\boldsymbol{\Sigma}}_k$, where is $\alpha \in \{1, 5, 10, \cdots, 200\}$ so that $\boldsymbol{n} \sim \mathcal{N}\big(\mathbf{0},\, \boldsymbol{\Sigma}_k\big)$.

Reference Empirical Baseline. For each experimental setting (noise level, acceleration rate, dataset) and image, we performed $N = 3,000$ MC trials (See Appendix G), each time adding simulated noise $\boldsymbol{n}$ drawn from $\boldsymbol{\Sigma}_k$. We reconstruct each noisy measurement and compute *empirical* variance maps by measuring voxelwise sample variance across all trials. The empirical variance maps serve as a gold standard reference.

Variance Estimation. Variance maps computed by our proposed *randomized Jacobian sketching* algorithm were compared to the empirical variance maps (our benchmark) to assess the accuracy of the proposed method for capturing voxel-wise noise propagation. The sketching matrix $\boldsymbol{V}_S$ had $S = 1000$ column probing vectors, which provides a balance between the estimation accuracy and computation (See Appendix F). Percent (%) Pearson Correlation Coefficient (PCC), and Normalized Root-Mean-Square Error (NRMSE) between computed and empirical variance maps were measured, and mean±std across all test slices were calculated to quantify the performance of our method. Significance of the distributional differences between voxelwise noise maps were assessed by using Wilcoxon signed-rank tests. Our code is available at https://github.com/onat-dalmaz/deep_recon_noise.

4.2. Generalizability Across Reconstruction Methods

To assess how well our variance estimator generalizes across different DL-based reconstruction methods, we fix $R = 8$ and $\boldsymbol{\Sigma}_k = \widetilde{\boldsymbol{\Sigma}}_k$, then train each model (knee and brain) and estimate NV. Table 1 shows that our method achieves near-perfect correlations to the gold standard, with no statistically significant difference betweeen noise distributions ($p > 0.05$). Although the U-Net yields slightly higher NRMSE (1.7%), these values remain small in absolute terms. One contributing factor to this might be U-Net's lack of iterative data-consistency, which applies a single-pass nonlinear denoising. This leads to both lower-fidelity reconstructions (See Table 9 in Appendix), and deviations from our method's linear approximation, resulting in minor local estimation errors.

Figures 1 and 6 (Appendix) illustrate representative noise-variance maps from knee and brain slices, respectively. Our estimator generally reproduces the intensity ranges and spatial patterns observed in the empirical reference, showing only minor deviations near regions of sharp intensity transitions, such as tissue boundaries or areas with rapid signal variation. These localized discrepancies remain negligible relative to the overall noise distribution and do not affect

Table 1. Mean±std PCC and NRMSE between noise-variance maps calculated by our method and by reference empirical baseline, across different DL reconstruction methods for knee and brain data at $R = 8$, $\alpha = 1$.

Method	Knee		Brain	
	PCC (%)	NRMSE (%)	PCC (%)	NRMSE (%)
E2E-VarNet	99.9 ± 0.0	0.7 ± 0.0	99.9 ± 0.0	0.5 ± 0.1
MoDL	99.9 ± 0.0	0.5 ± 0.0	99.7 ± 0.0	1.1 ± 0.1
U-Net	99.4 ± 0.0	1.7 ± 0.2	99.7 ± 0.0	1.8 ± 0.2
VORTEX	99.9 ± 0.0	0.6 ± 0.0	99.9 ± 0.0	0.6 ± 0.1
N2R	99.9 ± 0.0	0.6 ± 0.0	99.9 ± 0.0	0.7 ± 0.1
SSDU	99.9 ± 0.0	0.4 ± 0.0	99.8 ± 0.0	0.9 ± 0.1

Table 2. Computational and memory efficiency (per slice) on knee dataset for standard unrolled architecture, i.e. as used in E2E-VarNet, VORTEX, N2R, SSDU.

Metric	Empirical	Proposed	Naive
Time	54.0 s	1.3 s	880.2 s
Storage	3513.3 MB	1.1 MB	1.1 MB

the overall reliability of the framework. Across a variety of DL models, our estimator consistently provides robust noise estimates that align well with empirical simulations.

4.3. Computational and Memory Benefits

Table 2 compares the computational time and memory usage for three variance-estimation strategies for a single slice on the unrolled architecture from the knee dataset: *Empirical*, *Proposed*, and *Naive* (based on Eq. 11). *Empirical* simulations take about a minute and require storing multiple reconstructions. In contrast, *Naive* avoids large storage overhead, but its exhaustive row-by-row Jacobian calculation is significantly more time-consuming. In comparison, our *proposed* estimator balances the two ends: reducing the computation by over an order of magnitude relative to the empirical baseline by leveraging JVPs, while circumventing storage of reconstructed images. These results illustrate how the proposed algorithm attains substantial speed and memory advantages, making it a practical tool for routine use, especially on large datasets. See Appendix I for efficiency/scalability of our method with regards to network complexity and architecture, and Appendix J for a practical comparison with the empirical baseline at low trial counts.

4.4. Generalization to Various Undersampling Schemes

For the baseline noise level ($\alpha = 1$), we vary the acceleration factor $R \in \{4, 8, 12, 16, 24\}$, training a separate model for each R. Figures 7, 8 (Appendix) illustrate variance maps for representative knee and brain slices, respectively. Although higher acceleration factors R decrease the number of k-space samples—thus elevating ill-posedness and noise—our noise estimations continue to align closely with empirical references across all R, remaining their accuracy over a broad range of undersampling rates.

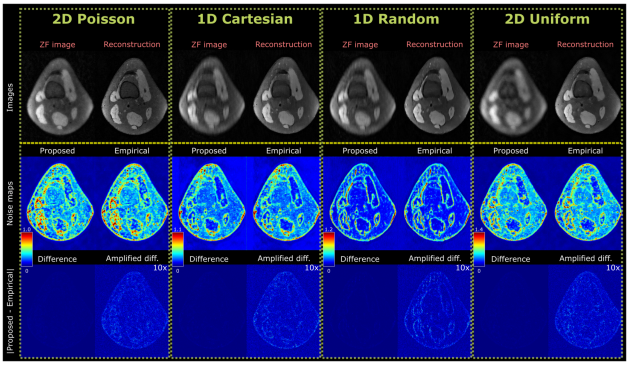

Figure 2. Each column corresponds to a different undersampling pattern (E2E-VarNet on knee data at $R = 8, \alpha = 1$). In each column: **(top row)** shows ZF and reconstructed images; **(middle row)** displays NV maps derived by the proposed method and empirical simulations; **(bottom row)** presents difference and amplified (x10) difference maps between the proposed and empirical variance maps to highlight spatial discrepancies. Color bars indicate each panel's relative noise map display window.

We also assessed the generalizability of our method under varying undersampling patterns, including 1D uniform Cartesian and 1D random Cartesian undersampling, and 2D uniformly random and Poisson-disc (default sampling pattern used in this study) undersampling (see Figure 10 in Appendix). For each pattern, we train a dedicated model, then estimate variance in representative slices. Figure 2 demonstrates that our noise maps retain strong agreement with empirical results under all sampling schemes—whether uniform or random. This consistency arises because the operator $\boldsymbol{A}$ inherently encodes the sampling mask; our method only relies on $\boldsymbol{A}$ as an implicit linear operator, making it *undersampling-agnostic*. These results demonstrate robustness to substantial changes in sampling pattern and acceleration rate, underscoring its versatility in diverse imaging scenarios.

4.5. Robustness to k-space Noise Level

We systematically increase the k-space noise magnitude by varying α as described in Section 4.1, thereby simulating growing covariance matrices (and thus SNR levels). Figures 3, 11 (Appendix) show variance maps for representative brain and knee slices from the proposed method and from empirical simulations, and Figure 13 (Appendix) shows the error between them as alpha grows. Although the reconstructed images become visibly noisier at higher α, networks still perform noticeable denoising. As α increases, empirical maps spread variability more uniformly across tissues, leading to a more homogeneous noise pattern without accentuating different tissue types. For moderate noise scaling ($\alpha < 50$), the proposed and empirical variance maps closely match. Note that extreme noise levels ($50\times$–$200\times$) correspond to SNR values well below 10–15 dB and thus far beyond what is considered clinically relevant in routine

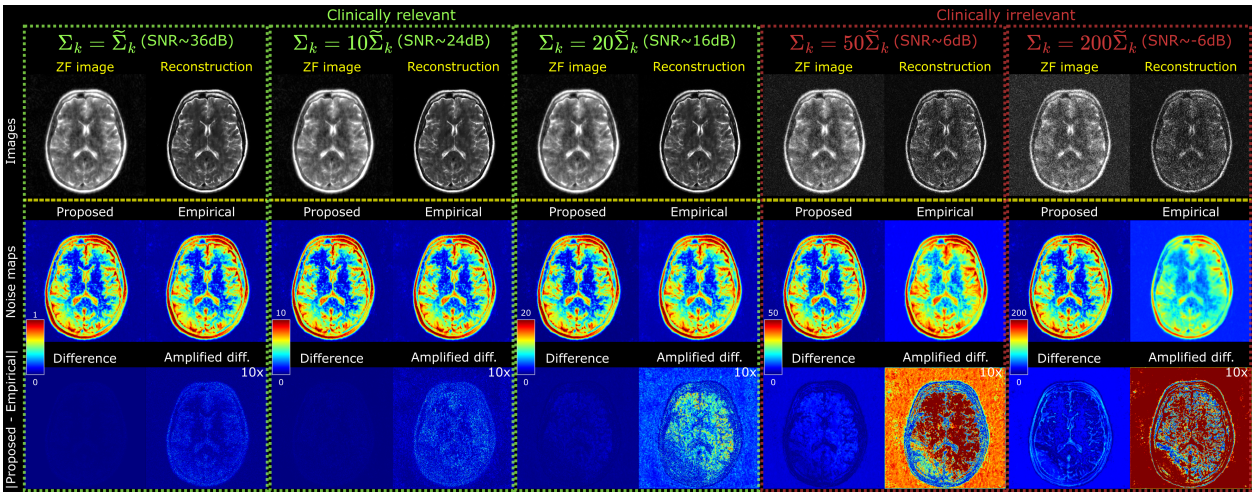

Figure 3. Each column corresponds to a different noise scaling factor α resulting in varying SNR scenarios (E2E-VarNet on knee data at $R = 8$). In each column: **(top row)** shows ZF and reconstructed images; **(middle row)** displays NV maps derived by the proposed method and empirical simulations; **(bottom row)** presents difference and amplified (x10) difference maps between the proposed and empirical variance maps to highlight spatial discrepancies. Color bars indicate each panel's relative noise map display window. Note that $\alpha = 50, 200$ corresponds to SNR values well under a 10dBs, which is often cited as a threshold for diagnostic utility (Brown et al., 2014).

practice (Brown et al., 2014; Westbrook et al., 2018). At these regimes, differences emerge in low-intensity tissue or around sharp edges, reflecting how the network's learned regularization systematically suppresses part of the injected noise compared to our estimation. Nevertheless, these difference maps remain small under practical SNR conditions, indicating that our approach effectively captures noise propagation for typical or moderately elevated noise. We provide additional analyses and discussions in Appendices L and M.

5. Discussion and Conclusion

Our localization at $\boldsymbol{x}_0$ presupposes that $\|\boldsymbol{A}^H\boldsymbol{n}\| \ll \|\boldsymbol{x}_0\|$—i.e., the noise perturbation in the image domain must not significantly deviate from the baseline signal, for the first-order Taylor approximation to remain valid. Clinically, typical MRI acquisitions usually have sufficiently high SNR for this condition to hold, so the linear approximation remains valid (Brown et al., 2014; Westbrook et al., 2018). However, once noise levels become so large that $\|\boldsymbol{A}^H\boldsymbol{n}\|$ approaches or exceeds $\|\boldsymbol{x}_0\|$, the nonlinearities in f can no longer be ignored. In such extreme cases—such as noise magnitudes exceeding the baseline signal—our method may under- or over-estimate variance, as the learned regularization could reshape these large perturbations in ways that can not be accounted for by a linear approximation. Despite these nuances at high noise regimes, the strong agreement between our method and empirical simulations under realistic acquisition conditions highlights the practical value and efficacy of our approach.

Noise distribution in fully-sampled or linear reconstructions tends to be fairly uniform, modulated by coil sensitivities (Robson et al., 2008). However, different DL reconstructions exhibit distinct noise profile for the same slice, shaped by factors such as regularization, optimization criteria, and the network's learned biases. Moreover noise distribution often depends on the underlying anatomy and k-space data, rather than being uniformly distributed. Indeed, inspection of the noise maps suggests that spatial noise amplification in DL reconstructions mirrors features in ZF inputs and final outputs. This localized noise amplification has important clinical implications. In routine MRI assessments, radiologists focus on specific regions of interest (ROI) such as cartilage in knee scans or pathology-affected areas in the brain, rather than examining the entire field of view uniformly (Rubenstein et al., 1997). Therefore, even if global metrics indicate high reconstruction quality, a locally elevated NV in diagnostically relevant ROIs could compromise clinical interpretation (Sijbers et al., 1998; Lerski et al., 1993). The ability to accurately quantify noise and preserve reliable diagnostic signal in crucial ROIs is pivotal for the successful clinical translation of DL-MRI methods (Kiryu et al., 2023). Our method efficiently provides accurate spatially resolved NV maps that highlight regions with elevated uncertainty, enabling radiologists to review diagnostically vulnerable areas and supporting better clinical decisions.

The variance estimation framework presented in this work naturally aligns with the broader concept of *uncertainty quantification* in deep neural networks and primarily captures *aleatoric* uncertainty, arising from measurement noise in k-space. By interpreting voxel-wise variance as an indicator of low SNR or ill-conditioned regions, this approach facilitates reliable confidence assessments in clinical settings, helping guide diagnostic decisions ("Is this subtle

lesion credible?") and acquisition strategies ("Should additional k-space samples be acquired here?"). Our proposed estimator efficiently computes voxel-wise uncertainty via JVPs, making it a promising approach for large deep MRI reconstruction pipelines to explicitly characterize noise and enhance interpretability. In the future, this method could be combined with *inherently stochastic reconstruction techniques*, i.e. diffusion models (Ozturkler et al., 2023; Chung & Ye, 2022), to also capture *epistemic* uncertainty arising from the model. While our current focus is on estimating the *diagonal* elements of the image covariance matrix (i.e., voxelwise variances), extending these techniques to capture *off-diagonal* terms also remains an important frontier. For instance, in functional MRI, cross-covariance structures could illuminate shared noise sources or functional connectivity across different brain regions (Varoquaux et al., 2010; Behrouz et al., 2024).

Many imaging modalities, such as CT and PET share similar challenges related to noise characterization, reconstruction under limited data, and estimating uncertainty in predictions (Wang et al., 2020). For instance, in CT and PET, where dose reduction is a key objective, quantifying uncertainty due to measurement noise could enhance confidence in low-dose reconstructions, improving clinical decision-making (Yang et al., 2018). While our primary goal in this study was to specifically address multi-coil accelerated MRI due to its complex acquisition model and the widespread adoption of DL-based reconstructions where explicit noise propagation has been relatively understudied, our framework is fundamentally generalizable. It is not limited solely to Gaussian noise assumptions and can be extended beyond MRI. The method requires only a known forward operator A describing the measurement process (e.g., X-ray projections in CT) and an appropriate noise covariance Σ_k reflecting the modality-specific noise statistics. For instance, extending our voxel-wise NV estimation approach to CT could involve incorporating Poisson-Gaussian noise models or variance-stabilizing transformations directly into Σ_k. Thus, while we consider such extensions an important future direction, the current manuscript lays the foundational theoretical and computational framework necessary to address these broader applications.

The principles developed in this study hold significant potential for broader image-to-image tasks in computer vision, including super-resolution, denoising, inpainting, or synthesis by appropriate modifications based on the input noise distribution and forward corruption model (Ledig et al., 2017; Zhang et al., 2017; Pathak et al., 2016; Dalmaz et al., 2022; Özbey et al., 2023). In these applications, accurately estimating the spatial distribution of uncertainty at the pixel level can improve interpretability and trustworthiness, especially in high-stakes scenarios such as medical diagnostics or autonomous systems (Kendall & Gal, 2017).

Voxel-wise NV information has been demonstrated to enhance downstream imaging tasks—such as denoising and segmentation—across multiple medical imaging modalities, including MRI, CT, and ultrasound. Classical adaptive filters like BM3D have been shown to significantly benefit from accurate NV information, enabling effective noise reduction without sacrificing critical diagnostic details (Hanchate & Joshi, 2020; Li et al., 2014). Similarly, adaptive filtering methods that incorporate spatially varying noise variance maps improve diagnostic quality in low-dose CT and ultrasound imaging scenarios (Hariharan et al., 2020; Yu & Acton, 2002). In contemporary DL, explicitly incorporating voxel-wise NV maps has consistently been demonstrated to reduce false positives and increase accuracy in segmentation and denoising networks (Dou et al., 2025; Wei et al., 2022). Thus, our proposed NV estimation framework not only enhances the interpretability of DL-based MRI reconstructions but also offers direct practical benefits by supporting robust performance improvements in critical downstream clinical tasks.

In conclusion, we introduced a model-agnostic Jacobian sketching algorithm for estimating voxel-wise NV in DL MRI reconstructions. Our approach efficiently and accurately captures how acquisition noise propagates through nonlinear, multi-coil reconstructions without costly MC simulations. Extensive experiments demonstrate strong alignment with reference variance maps across diverse architectures, suggesting that our method can be readily incorporated into existing deep reconstruction pipelines. Beyond improving interpretability of DL reconstructions, these noise estimates hold promise for guiding adaptive acquisition, assessing image quality, and facilitating clinically actionable uncertainty quantification.

Acknowledgements

Authors would like to thank Moritz Alexander Bolling for helpful discussions on first order methods and statistical tests. This work was supported by the following grants: NIH R01AR077604, R0EB009055, R01CA249893, R01EB002524.

Impact Statement

This work aims to enhance the reliability and interpretability of deep learning-based MRI reconstruction methods, potentially improving clinical decision-making by providing more accurate and trustworthy imaging tools. While its primary objective is to advance the field of machine learning, we acknowledge that such advancements may have broader societal implications, including increased accessibility to high-quality diagnostic imaging and more efficient healthcare workflows. At present, we see no specific negative ethical or societal issues that warrant particular attention.

References

Adamson, P. M., Gunel, B., Dominic, J., Desai, A. D., Spielman, D., Vasanawala, S., Pauly, J. M., and Chaudhari, A. SSFD: Self-supervised feature distance as an MR image reconstruction quality metric. In *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, 2021. URL https://openreview.net/forum?id=dgMvTzf6M_3.

Adamson, P. M., Desai, A. D., Dominic, J., Bluethgen, C., Wood, J. P., Syed, A. B., Boutin, R. D., Stevens, K. J., Vasanawala, S., Pauly, J. M., Chaudhari, A. S., and Gunel, B. Using deep feature distances for evaluating MR image reconstruction quality. In *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*, 2023. URL https://openreview.net/forum?id=AUiZyqYiGb.

Aggarwal, H. K., Mani, M. P., and Jacob, M. Modl: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38:394–405, 2 2019. ISSN 1558254X. doi: 10.1109/TMI.2018.2865356.

Aja-Fernández, S., Vegas-Sánchez-Ferrero, G., and Tristán-Vega, A. Noise estimation in parallel mri: Grappa and sense. *Magnetic Resonance Imaging*, 32:281–290, 4 2014. ISSN 0730725X. doi: 10.1016/j.mri.2013.12.001.

Akcakaya, M., Basha, T. A., Manning, W. J., and Nezafat, R. Efficient calculation of g-factors for cg-sense in high dimensions: noise amplification in random undersampling. *Journal of Cardiovascular Magnetic Resonance*, 16, 1 2014. ISSN 1532429X. doi: 10.1186/1532-429X-16-S1-W28.

Akçakaya, M., Moeller, S., Weingärtner, S., and Uğurbil, K. Scan-specific robust artificial-neural-networks for k-space interpolation (raki) reconstruction: Database-free deep learning for fast imaging. *Magnetic Resonance in Medicine*, 81:439–453, 1 2019. ISSN 15222594. doi: 10.1002/mrm.27420.

Alkan, C., Mardani, M., Liao, C., Li, Z., Vasanawala, S. S., and Pauly, J. M. Autosamp: Autoencoding k-space sampling via variational information maximization for 3d mri. *IEEE Transactions on Medical Imaging*, 2024. ISSN 1558254X. doi: 10.1109/TMI.2024.3443292.

Behrouz, A., Delavari, P., and Hashemi, F. Unsupervised representation learning of brain activity via bridging voxel activity and functional connectivity. In Salakhutdinov, R., Kolter, Z., Heller, K., Weller, A., Oliver, N., Scarlett, J., and Berkenkamp, F. (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 3347–3381. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/behrouz24a.html.

Bekas, C., Kokiopoulou, E., and Saad, Y. An estimator for the diagonal of a matrix. *Applied Numerical Mathematics*, 57(11):1214–1229, 2007. ISSN 0168-9274. doi: https://doi.org/10.1016/j.apnum.2007.01.003. URL https://www.sciencedirect.com/science/article/pii/S0168927407000244. Numerical Algorithms, Parallelism and Applications (2).

Breuer, F. A., Kannengiesser, S. A., Blaimer, M., Seiberlich, N., Jakob, P. M., and Griswold, M. A. General formulation for quantitative g-factor calculation in grappa reconstructions. *Magnetic Resonance in Medicine*, 62:739–746, 2009. ISSN 15222594. doi: 10.1002/mrm.22066.

Brown, R., Cheng, Y., Haacke, E., Thompson, M., and Venkatesan, R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley, 2014. ISBN 9781118633977. URL https://books.google.com/books?id=rQGCAwAAQBAJ.

Chaudhari, A. S., Sandino, C. M., Cole, E. K., Larson, D. B., Gold, G. E., Vasanawala, S. S., Lungren, M. P., Hargreaves, B. A., and Langlotz, C. P. Prospective deployment of deep learning in mri: A framework for important considerations, challenges, and recommendations for best practices. *Journal of Magnetic Resonance Imaging*, 54(2):357–371, 2021. doi: https://doi.org/10.1002/jmri.27331. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.27331.

Chen, Y., Schönlieb, C.-B., Liò, P., Leiner, T., Dragotti, P. L., Wang, G., Rueckert, D., Firmin, D., and Yang, G. Ai-based reconstruction for fast mri—a systematic review and meta-analysis. *Proceedings of the IEEE*, 110 (2):224–245, 2022. doi: 10.1109/JPROC.2022.3141367.

Chung, H. and Ye, J. C. Score-based diffusion models for accelerated mri. *Medical Image Analysis*, 80, 8 2022. ISSN 13618423. doi: 10.1016/j.media.2022.102479.

Dalmaz, O., Yurt, M., and Çukur, T. Resvit: Residual vision transformers for multimodal medical image synthesis. *IEEE Transactions on Medical Imaging*, 41(10):2598–2614, 2022. doi: 10.1109/TMI.2022.3167808.

Dalmaz, O., Desai, A., Chaudhari, A., and Hargreaves, B. A. Noise-induced variability quantification in deep learning-based mri reconstructions. In *Proceedings of the 32nd Annual Meeting of ISMRM*, pp. 3–9, Singapore, May 2024.

Darestani, M. Z., Chaudhari, A. S., and Heckel, R. Measuring robustness in deep learning based compressive sensing. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 2433–2444. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/darestani21a.html.

Dawood, P., Breuer, F., Homolya, I., Stebani, J., Gram, M., Jakob, P. M., Zaiss, M., and Blaimer, M. A novel image space formalism of fourier domain interpolation neural networks for noise propagation analysis, 2024. URL https://arxiv.org/abs/2402.17410.

Desai, A. D., Gunel, B., Ozturkler, B. M., Beg, H., Vasanawala, S., Hargreaves, B. A., Ré, C., Pauly, J. M., and Chaudhari, A. S. Vortex: Physics-driven data augmentations using consistency training for robust accelerated mri reconstruction. In Konukoglu, E., Menze, B., Venkataraman, A., Baumgartner, C., Dou, Q., and Albarqouni, S. (eds.), *PMLR*, volume 172 of *Proceedings of Machine Learning Research*, pp. 325–352. PMLR, 06–08 Jul 2022a. URL https://proceedings.mlr.press/v172/desai22a.html.

Desai, A. D., Schmidt, A. M., Rubin, E. B., Sandino, C. M., Black, M. S., Mazzoli, V., Stevens, K. J., Boutin, R. D., Ré, C., Gold, G. E., Hargreaves, B. A., and Chaudhari, A. S. Skm-tea: A dataset for accelerated mri reconstruction with dense image labels for quantitative clinical evaluation. *ArXiv*, abs/2203.06823, 2022b. URL https://api.semanticscholar.org/CorpusID:244907741.

Desai, A. D., Ozturkler, B. M., Sandino, C. M., Boutin, R., Willis, M., Vasanawala, S., Hargreaves, B. A., Ré, C., Pauly, J. M., and Chaudhari, A. S. Noise2recon: Enabling snr-robust mri reconstruction with semi-supervised and self-supervised learning. *Magnetic Resonance in Medicine*, 90(5):2052–2070, 2023. doi: https://doi.org/10.1002/mrm.29759. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.29759.

Dietrich, O., Raya, J. G., Reeder, S. B., Reiser, M. F., and Schoenberg, S. O. Measurement of signal-to-noise ratios in mr images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *Journal of Magnetic Resonance Imaging*, 26:375–385, 8 2007. ISSN 10531807. doi: 10.1002/jmri.20969.

Dou, Q., Wang, Z., Feng, X., Campbell-Washburn, A. E., Mugler III, J. P., and Meyer, C. H. Mri denoising with a non-blind deep complex-valued convolutional neural network. *NMR in Biomedicine*, 38(1):e5291, 2025. doi: https://doi.org/10.1002/nbm.5291. URL https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/nbm.5291.

Edupuganti, V., Mardani, M., Vasanawala, S., and Pauly, J. Uncertainty quantification in deep mri reconstruction. *IEEE Transactions on Medical Imaging*, 40(1):239–250, 2021. doi: 10.1109/TMI.2020.3025065.

Fabian, Z., Tinaz, B., and Soltanolkotabi, M. HUMUS-net: Hybrid unrolled multi-scale network architecture for accelerated MRI reconstruction. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=z0M3qHDqH20.

Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Gower, R. M., Richtárik, P., and Bach, F. Stochastic quasi-gradient methods: variance reduction via jacobian sketching. *Math. Program.*, 188(1):135–192, July 2021. ISSN 0025-5610. doi: 10.1007/s10107-020-01506-0. URL https://doi.org/10.1007/s10107-020-01506-0.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., and Haase, A. Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic Resonance in Medicine*, 47:1202–1210, 2002. ISSN 07403194. doi: 10.1002/mrm.10171.

Hammernik, K., Küstner, T., Yaman, B., Huang, Z., Rueckert, D., Knoll, F., and Akçakaya, M. Physics-driven deep learning for computational magnetic resonance imaging:

Combining physics and machine learning for improved medical imaging. *IEEE Signal Processing Magazine*, 40 (1):98–114, 2023. doi: 10.1109/MSP.2022.3215288.

Hanchate, V. and Joshi, K. Mri denoising using bm3d equipped with noise invalidation denoising technique and vst for improved contrast. *SN Applied Sciences*, 2(2):234, 2020. doi: 10.1007/s42452-020-1937-7. URL https://doi.org/10.1007/s42452-020-1937-7.

Hariharan, S. G., Strobel, N., Kaethner, C., Kowarschik, M., Fahrig, R., and Navab, N. Data-driven estimation of noise variance stabilization parameters for low-dose x-ray images. *Physics in Medicine Biology*, 65(22):225027, nov 2020. doi: 10.1088/1361-6560/abbc82. URL https://dx.doi.org/10.1088/1361-6560/abbc82.

Heckel, R. Provable continual learning via sketched jacobian approximations. In Camps-Valls, G., Ruiz, F. J. R., and Valera, I. (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 10448–10470. PMLR, 28–30 Mar 2022. URL https://proceedings.mlr.press/v151/heckel22a.html.

Heckel, R., Jacob, M., Chaudhari, A., Perlman, O., and Shimron, E. Deep learning for accelerated and robust mri reconstruction. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 37(3):335–368, 2024. doi: 10.1007/s10334-024-01173-8. URL https://doi.org/10.1007/s10334-024-01173-8.

Hoppe, F., Verdun, C. M., Laus, H., Endt, S., Menzel, M. I., Krahmer, F., and Rauhut, H. Imaging with;confidence: Uncertainty quantification for;high-dimensional undersampled mr images. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part LXXVIII*, pp. 432–450, Berlin, Heidelberg, 2024. Springer-Verlag. ISBN 978-3-031-73228-7. doi: 10.1007/978-3-031-73229-4_25. URL https://doi.org/10.1007/978-3-031-73229-4_25.

Hutchinson, M. F. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *J. Commun. Statist. Simula.*, 19(2):433–450, 1990. URL https://cir.nii.ac.jp/crid/1573668925763299072.

Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A. G., and Tamir, J. Robust compressed sensing mri with deep generative priors. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 14938–14954. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/7d6044e95a16761171b130dcb476a43e-Paper.pdf.

Johnson, P. M., Lin, D. J., Zbontar, J., Zitnick, C. L., Sriram, A., Muckley, M., Babb, J. S., Kline, M., Ciavarra, G., Alaia, E., Samim, M., Walter, W. R., Calderon, L., Pock, T., Sodickson, D. K., Recht, M. P., and Knoll, F. Deep learning reconstruction enables prospectively accelerated clinical knee mri. *Radiology*, 307(2):e220425, 2023. doi: 10.1148/radiol.220425. URL https://doi.org/10.1148/radiol.220425. PMID: 36648347.

Jun, Y., Shin, H., Eo, T., and Hwang, D. Joint deep model-based mr image and coil sensitivity reconstruction network (joint-icnet) for fast mri. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5266–5275, 2021. doi: 10.1109/CVPR46437.2021.00523.

Kastryulin, S., Zakirov, J., Pezzotti, N., and Dylov, D. V. Image quality assessment for magnetic resonance imaging. *IEEE Access*, 11:14154–14168, 2023. doi: 10.1109/ACCESS.2023.3243466.

Kellman, P. and McVeigh, E. R. Image reconstruction in snr units: A general method for snr measurement. *Magnetic Resonance in Medicine*, 54:1439–1447, 2005. ISSN 07403194. doi: 10.1002/mrm.20713.

Kendall, A. and Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5574–5584, 2017.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

Kiryu, S., Akai, H., Yasaka, K., Tajima, T., Kunimatsu, A., Yoshioka, N., Akahane, M., Abe, O., and Ohtomo, K. Clinical impact of deep learning reconstruction in mri. *RadioGraphics*, 43(6):e220133, 2023. doi: 10.1148/rg.220133. URL https://doi.org/10.1148/rg.220133. PMID: 37200221.

Knoll, F., Hammernik, K., Kobler, E., Pock, T., Recht, M. P., and Sodickson, D. K. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic Resonance in Medicine*, 81:116–128, 11 2019. ISSN 15222594. doi: 10.1002/mrm.27355.

Knoll, F., Hammernik, K., Zhang, C., Moeller, S., Pock, T., Sodickson, D. K., and Akcakaya, M. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends,

and issues. *IEEE Signal Processing Magazine*, 37:128–140, 1 2020a. ISSN 15580792. doi: 10.1109/MSP.2019.2950640.

Knoll, F., Murrell, T., Sriram, A., Yakubova, N., Zbontar, J., Rabbat, M., Defazio, A., Muckley, M. J., Sodickson, D. K., Zitnick, C. L., and Recht, M. P. 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, 2020b. doi: https://doi.org/10.1002/mrm.28338. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28338.

Knoll, F., Zbontar, J., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., Parente, M., Geras, K. J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzalv, M., Romero, A., Rabbat, M., Vincent, P., Pinkerton, J., Wang, D., Yakubova, N., Owens, E., Zitnick, C. L., Recht, M. P., Sodickson, D. K., and Lui, Y. W. fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, 2020c. doi: 10.1148/ryai.2020190007. URL https://doi.org/10.1148/ryai.2020190007. PMID: 32076662.

Küstner, T., Hammernik, K., Rueckert, D., Hepp, T., and Gatidis, S. Predictive uncertainty in deep learning–based mr image reconstruction using deep ensembles: Evaluation on the fastmri data set. *Magnetic Resonance in Medicine*, 92(1):289–302, 2024. doi: https://doi.org/10.1002/mrm.30030. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.30030.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., and Shi, W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690, 2017. doi: 10.1109/CVPR.2017.19.

Lerski, R., Straughan, K., Schad, L., Boyce, D., Blüml, S., and Zuna, I. Viii. mr image texture analysis—an approach to tissue characterization. *Magnetic Resonance Imaging*, 11(6):873–887, 1993. ISSN 0730-725X. doi: https://doi.org/10.1016/0730-725X(93)90205-R. URL https://www.sciencedirect.com/science/article/pii/0730725X9390205R.

Li, H., Krishnan, A., Wu, J., Kolouri, S., Pilly, P. K., and Braverman, V. Lifelong learning with sketched structural regularization. In Balasubramanian, V. N. and Tsang, I. (eds.), *Proceedings of The 13th Asian Conference on Machine Learning*, volume 157 of *Proceedings of Machine Learning Research*, pp. 985–1000. PMLR, 17–19 Nov 2021. URL https://proceedings.mlr.press/v157/li21b.html.

Li, Z., Yu, L., Trzasko, J. D., Lake, D. S., Blezek, D. J., Fletcher, J. G., McCollough, C. H., and Manduca, A. Adaptive nonlocal means filtering based on local noise level for ct denoising. *Medical Physics*, 41(1):011908, 2014. doi: https://doi.org/10.1118/1.4851635. URL https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4851635.

Liang, D., Cheng, J., Ke, Z., and Ying, L. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal Processing Magazine*, 37 (1):141–151, 2020. doi: 10.1109/MSP.2019.2950557.

Macovski, A. Noise in mri. *Magnetic Resonance in Medicine*, 36(3):494–497, 1996. doi: https://doi.org/10.1002/mrm.1910360327. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.1910360327.

Mardani, M., Gong, E., Cheng, J. Y., Vasanawala, S. S., Zaharchuk, G., Xing, L., and Pauly, J. M. Deep generative adversarial neural networks for compressive sensing mri. *IEEE Transactions on Medical Imaging*, 38(1):167–179, 2019. doi: 10.1109/TMI.2018.2858752.

Mason, A., Rioux, J., Clarke, S. E., Costa, A., Schmidt, M., Keough, V., Huynh, T., and Beyea, S. Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of mr images. *IEEE Transactions on Medical Imaging*, 39(4):1064–1072, 2020. doi: 10.1109/TMI.2019.2930338.

Muckley, M. J., Riemenschneider, B., Radmanesh, A., Kim, S., Jeong, G., Ko, J., Jun, Y., Shin, H., Hwang, D., Mostapha, M., Arberet, S., Nickel, D., Ramzi, Z., Ciuciu, P., Starck, J. L., Teuwen, J., Karkalousos, D., Zhang, C., Sriram, A., Huang, Z., Yakubova, N., Lui, Y. W., and Knoll, F. Results of the 2020 fastmri challenge for machine learning mr image reconstruction, 9 2021. ISSN 1558254X.

Narnhofer, D., Effland, A., Kobler, E., Hammernik, K., Knoll, F., and Pock, T. Bayesian uncertainty estimation of learned variational mri reconstruction. *IEEE Transactions on Medical Imaging*, 41:279–291, 2021. URL https://api.semanticscholar.org/CorpusID:231918509.

Ozturkler, B., Liu, C., Eckart, B., Mardani, M., Song, J., and Kautz, J. Smrd: Sure-based robust mri reconstruction with diffusion models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 14222 LNCS, pp. 199–209. Springer Science

and Business Media Deutschland GmbH, 2023. ISBN 9783031438974. doi: 10.1007/978-3-031-43898-1_20.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. *PyTorch: an imperative style, high-performance deep learning library*. Curran Associates Inc., Red Hook, NY, USA, 2019.

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., and Efros, A. A. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536–2544, 2016. doi: 10.1109/CVPR.2016.278.

Peng, W., Feng, L., Zhao, G., and Liu, F. Learning optimal k-space acquisition and reconstruction using physics-informed neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20794–20803, June 2022.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. Sense: Sensitivity encoding for fast mri. *Magnetic Resonance in Medicine*, 42:952–962, 1999. ISSN 07403194. doi: 10.1002/(SICI)1522-2594(199911)42:5⟨952::AID-MRM16⟩3.0.CO;2-S.

Robson, P. M., Grant, A. K., Madhuranthakam, A. J., Lattanzi, R., Sodickson, D. K., and McKenzie, C. A. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magnetic Resonance in Medicine*, 60 (4):895–907, 2008. doi: https://doi.org/10.1002/mrm.21728. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.21728.

Roemer, P. B., Edelstein, W. A., Hayes, C. E., Souza, S. P., and Mueller, O. M. The nmr phased array. *Magnetic Resonance in Medicine*, 16(2):192–225, 1990. doi: https://doi.org/10.1002/mrm.1910160203. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.1910160203.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F. (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Roosta-Khorasani, F. and Ascher, U. M. Improved bounds on sample size for implicit matrix trace estimators. *Foundations of Computational Mathematics*, 15:1187–1212, 2013. URL https://api.semanticscholar.org/CorpusID:2094935.

Rubenstein, J. D., Li, J. G., Majumdar, S., and Henkelman, R. M. Image resolution and signal-to-noise ratio requirements for mr imaging of degenerative cartilage. *American Journal of Roentgenology*, 169(4): 1089–1096, 1997. doi: 10.2214/ajr.169.4.9308470. URL https://doi.org/10.2214/ajr.169.4.9308470. PMID: 9308470.

Sijbers, J., den Dekker, A., Van Audekerke, J., Verhoye, M., and Van Dyck, D. Estimation of the noise in magnitude mr images. *Magnetic Resonance Imaging*, 16(1):87–90, 1998. ISSN 0730-725X. doi: https://doi.org/10.1016/S0730-725X(97)00199-9. URL https://www.sciencedirect.com/science/article/pii/S0730725X97001999.

Sodickson, D. K., Griswold, M. A., Jakob, P. M., Edelman, R. R., and Manning, W. J. Signal-to-noise ratio and signal-to-noise efficiency in smash imaging. *Magnetic Resonance in Medicine*, 41:1009–1022, 1999. ISSN 07403194. doi: 10.1002/(SICI)1522-2594(199905)41:5⟨1009::AID-MRM21⟩3.0.CO;2-4.

Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C. L., Yakubova, N., Knoll, F., and Johnson, P. End-to-end variational networks for accelerated mri reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II*, pp. 64–73, Berlin, Heidelberg, 2020a. Springer-Verlag. ISBN 978-3-030-59712-2. doi: 10.1007/978-3-030-59713-9_7. URL https://doi.org/10.1007/978-3-030-59713-9_7.

Sriram, A., Zbontar, J., Murrell, T., Zitnick, C. L., Defazio, A., and Sodickson, D. K. Grappanet: Combining parallel imaging with deep learning for multi-coil mri reconstruction. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14303–14310, 2020b. doi: 10.1109/CVPR42600.2020.01432.

Tezcan, K. C., Baumgartner, C. F., and Konukoglu, E. Sampling possible reconstructions of undersampled acquisitions in mr imaging with a deep learned prior. *IEEE Transactions on Medical Imaging*, 41:1885–1896, 2020. URL https://api.semanticscholar.org/CorpusID:222090823.

Thunberg, P. and Zetterberg, P. Noise distribution in sense- and grappa-reconstructed images: a computer simulation study. *Magnetic Resonance Imaging*, 25:1089–1094, 9 2007. ISSN 0730725X. doi: 10.1016/j.mri.2006.11.003.

Varoquaux, G., Gramfort, A., Poline, J.-B., and Thirion, B. Brain covariance selection: better individual functional connectivity models using population prior. In

Proceedings of the 24th International Conference on Neural Information Processing Systems - Volume 2, NIPS'10, pp. 2334–2342, Red Hook, NY, USA, 2010. Curran Associates Inc.

Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Wang, C., Guo, L., Wang, Y., Cheng, H., Yu, Y., and Wen, B. Progressive divide-and-conquer via subsampling decomposition for accelerated mri. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25128–25137, June 2024.

Wang, G., Ye, J. C., and De Man, B. Deep learning for tomographic image reconstruction. *Nature Machine Intelligence*, 2:737–748, 12 2020. doi: 10.1038/s42256-020-00273-z.

Wei, T., Wang, X., Li, X., and Zhu, S. Fuzzy subspace clustering noisy image segmentation algorithm with adaptive local variance non-local information and mean membership linking. *Engineering Applications of Artificial Intelligence*, 110:104672, 2022. ISSN 0952-1976. doi: https://doi.org/10.1016/j.engappai.2022.104672. URL https://www.sciencedirect.com/science/article/pii/S0952197622000069.

Wen, J., Ahmad, R., and Schniter, P. Task-driven uncertainty quantification innbsp;inverse problems vianbsp;conformal prediction. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part LX*, pp. 182–199, Berlin, Heidelberg, 2024. Springer-Verlag. ISBN 978-3-031-73026-9. doi: 10.1007/978-3-031-73027-6_11. URL https://doi.org/10.1007/978-3-031-73027-6_11.

Westbrook, C., Roth, C. K., and Talbot, J. *MRI in Practice*. Wiley-Blackwell, 5th edition, 2018. ISBN 978-1119391968.

Xiang, T., Yurt, M., Syed, A. B., Setsompop, K., and Chaudhari, A. DDM^2: Self-supervised diffusion MRI denoising with generative diffusion models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=0vqjc50HfcC.

Yaman, B., Hosseini, S. A. H., Moeller, S., Ellermann, J., Uğurbil, K., and Akçakaya, M. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84:3172–3191, 12 2020. ISSN 15222594. doi: 10.1002/mrm.28378.

Yaman, B., Hosseini, S. A. H., and Akcakaya, M. Zero-shot self-supervised learning for MRI reconstruction. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=085y6YPaYjP.

Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L., and Wang, G. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, 37(6):1348–1357, 2018. doi: 10.1109/TMI.2018.2827462.

Ying, L. and Sheng, J. Joint image reconstruction and sensitivity estimation in sense (jsense). *Magnetic Resonance in Medicine*, 57:1196–1202, 2007. ISSN 15222594. doi: 10.1002/mrm.21245.

Yu, Y. and Acton, S. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11(11):1260–1270, 2002. doi: 10.1109/TIP.2002.804276.

Yurt, M., Alkan, C., Schauman, S., Cao, X., Iyer, S., Liao, C., Cukur, T., Vasanawala, S., Pauly, J., and Setsompop, K. Semi-supervision for clinical contrast synthesis from magnetic resonance fingerprinting. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), MedNeurIPS Workshop*, 2022.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. doi: 10.1109/TIP.2017.2662206.

Özbey, M., Dalmaz, O., Dar, S. U. H., Bedel, H. A., Özturk, , Güngör, A., and Çukur, T. Unsupervised medical image translation with adversarial diffusion models. *IEEE Transactions on Medical Imaging*, 42(12):3524–3539, 2023. doi: 10.1109/TMI.2023.3290149.

A. Detailed Noise Model for Multi-Coil MRI

A.1. Coil Noise as Linear Combinations of Independent Sources

We assume that noise in each coil γ arises from an ensemble of zero-mean independent Gaussian sources $\{\xi_\tau\}$, each with standard deviation σ_τ (Macovski, 1996). Concretely,

$$\eta_\gamma(t) \;=\; \sum_\tau \omega_{(\gamma,\tau)}\, \xi_\tau(t). \tag{13}$$

Because each ξ_τ is Gaussian, $\eta_\gamma(t)$ must also be Gaussian. The weight $\omega_{(\gamma,\tau)}$ indicates how strongly the τ-th noise source couples to coil γ.

A.2. Coil Covariance Matrix $\widetilde{\mathbf{\Sigma}}_k$

Summarizing these noise contributions yields an $n_c \times n_c$ *coil covariance matrix* $\widetilde{\mathbf{\Sigma}}_k$, where

$$(\widetilde{\mathbf{\Sigma}}_k)_{\gamma,\gamma'} \;=\; \sum_\tau \sigma_\tau^2\, \omega_{(\gamma,\tau)}\, \big(\omega_{(\gamma',\tau)}\big)^*. \tag{14}$$

Here, $(\widetilde{\mathbf{\Sigma}}_k)_{\gamma,\gamma}$ is the NV in coil γ, and off-diagonal elements encode coil-to-coil covariances, and n_c is the number of channels in the imaging system. Since noise originates from physical sources that induce real-valued fluctuations, the cross-correlations between coils must satisfy Hermitian symmetry due to Maxwellian reciprocity in electromagnetic induction (Roemer et al., 1990; Brown et al., 2014). Consequently, $\widetilde{\mathbf{\Sigma}}_k$ is *Hermitian* (γ, γ' swap with a conjugate to see mathematically) and *positive semidefinite* by construction (as all covariance matrices), since it can be expressed in the form

$$\mathbf{W}\,\mathbf{D}\,\mathbf{W}^H,$$

where $\mathbf{W}$ contains $\omega_{(\gamma,\tau)}$, and $\mathbf{D}$ is diagonal of $\{\sigma_\tau^2\}$. Thus, for any vector $\boldsymbol{v} \in \mathbb{C}^{n_c}$,

$$\boldsymbol{v}^H\, \widetilde{\mathbf{\Sigma}}_k\, \boldsymbol{v} \;=\; \|\mathbf{D}^{1/2}\, \mathbf{W}^H\, \boldsymbol{v}\|^2 \;\geq\; 0,$$

confirming positive semidefiniteness.

A.3. Block-Diagonal Sample Covariance $\mathbf{\Sigma}_k$

Suppose noise is uncorrelated across n_f distinct k-space frequency locations, while retaining a coil-to-coil covariance $\widetilde{\mathbf{\Sigma}}_k \in \mathbb{C}^{n_c \times n_c}$ at each frequency (where n_c is the number of coils). If $\boldsymbol{n} \in \mathbb{C}^m$, with $m = n_f \cdot n_c$, denotes the stacked multi-coil noise over n_f frequency bins, then the *sample covariance matrix* of $\boldsymbol{n}$ can be expressed as a block diagonal:

$$\mathbf{\Sigma}_k \;=\; \underbrace{\mathrm{diag}\Big(\widetilde{\mathbf{\Sigma}}_k,\; \ldots,\; \widetilde{\mathbf{\Sigma}}_k\Big)}_{n_f \text{ blocks}} \;\in\; \mathbb{C}^{m \times m}.$$

Equivalently, we may write

$$\mathbf{\Sigma}_k \;=\; \boldsymbol{I}_{n_f} \,\otimes\, \widetilde{\mathbf{\Sigma}}_k,$$

where $\boldsymbol{I}_{n_f}$ is the $n_f \times n_f$ identity matrix and $\otimes$ denotes the Kronecker product. Each $n_c \times n_c$ diagonal block $\widetilde{\mathbf{\Sigma}}_k$ captures the coil-to-coil noise correlation structure for one frequency index, and repeating these blocks n_f times encodes independence across frequencies.

In practice, one can exploit a *compact* representation in which $\widetilde{\mathbf{\Sigma}}_k$ is used directly during linear-algebraic factorizations or operator-based manipulations, rather than forming the full $m \times m$ block diagonal. Such a compact approach reduces both memory and computational overhead significantly, as the repeated blocks need not be explicitly stored or inverted. Consequently, we directly utilize the coil covariance matrix $\widetilde{\mathbf{\Sigma}}_k$ in our computational framework for efficiency. Nevertheless, for theoretical analysis and proofs carried out in this study, it is more convenient to leverage the sample covariance matrix $\mathbf{\Sigma}_k$, thereby making explicit how noise correlations factor into the imaging operator $\boldsymbol{A}$ and the subsequent derivations. We now show that if $\widetilde{\mathbf{\Sigma}}_k$ is Hermitian positive semidefinite (HPSD), then $\mathbf{\Sigma}_k$ inherits these properties.

Hermitian. A block-diagonal matrix is Hermitian if and only if each diagonal block is Hermitian. Since $\widetilde{\mathbf{\Sigma}}_k^H = \widetilde{\mathbf{\Sigma}}_k$, we have

$$\mathbf{\Sigma}_k^H = (\boldsymbol{I}_{n_f} \otimes \widetilde{\mathbf{\Sigma}}_k)^H = \boldsymbol{I}_{n_f}^H \otimes \widetilde{\mathbf{\Sigma}}_k^H = \boldsymbol{I}_{n_f} \otimes \widetilde{\mathbf{\Sigma}}_k = \mathbf{\Sigma}_k.$$

Hence, $\mathbf{\Sigma}_k$ is Hermitian.

Positive Semidefinite (PSD). A Hermitian $\mathbf{\Sigma}_k$ is PSD if and only if $\mathbf{x}^H \mathbf{\Sigma}_k \mathbf{x} \geq 0$ for every $\mathbf{x} \in \mathbb{C}^m$. Partition $\mathbf{x}$ into n_f sub-vectors, $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_{n_f})$, each $\mathbf{x}_i \in \mathbb{C}^{n_c}$. Because $\mathbf{\Sigma}_k$ is block diagonal, we get

$$\mathbf{x}^H \, \mathbf{\Sigma}_k \, \mathbf{x} = \sum_{i=1}^{n_f} \Big(\mathbf{x}_i^H \, \widetilde{\mathbf{\Sigma}}_k \, \mathbf{x}_i \Big).$$

If $\widetilde{\mathbf{\Sigma}}_k$ is PSD, then each $\mathbf{x}_i^H \, \widetilde{\mathbf{\Sigma}}_k \, \mathbf{x}_i \geq 0$. Summing these terms yields a nonnegative result, thus $\mathbf{\Sigma}_k$ is PSD.

B. Existence of the Network Jacobian

Unrolled MRI networks alternate between *regularization* and *data-consistency* updates for K iterations, yielding a final reconstruction $\boldsymbol{x}^{(K)}$ (Hammernik et al., 2023; Liang et al., 2020). Concretely, each iteration has two steps:

1. **Regularization:**

$$\hat{\boldsymbol{x}}^{(k)} \;=\; \mathcal{F}^{(k)}\Big(\boldsymbol{x}^{(k-1)}\Big),$$

   where $\mathcal{F}^{(k)}(\cdot)$ is a *neural network* or learned operator (e.g., a CNN block) that acts as a trainable regularizer. Neural networks are well-known to be differentiable with respect to their inputs, assuming standard operations (convolution, ReLU, batch norm, etc.) (Goodfellow et al., 2016). Hence, the map $\boldsymbol{x}^{(k-1)} \mapsto \hat{\boldsymbol{x}}^{(k)}$ is differentiable.

2. **Data Consistency (DC):**

$$\boldsymbol{x}^{(k)} \;=\; \hat{\boldsymbol{x}}^{(k)} \;-\; \boldsymbol{A}^H\Big(\boldsymbol{A}\,\hat{\boldsymbol{x}}^{(k)} - \boldsymbol{y}\Big).$$

   Here, $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is the linear imaging operator, and $\boldsymbol{y} \in \mathbb{C}^m$ is the measured data. Since matrix-vector multiplication and addition/subtraction are *linear* (and thus differentiable) operations, the map $\hat{\boldsymbol{x}}^{(k)} \mapsto \boldsymbol{x}^{(k)}$ is also differentiable.

Composing Differentiable Steps. Because each iteration *composes* two differentiable transformations, the combined k-th iteration step

$$\boldsymbol{x}^{(k-1)} \;\mapsto\; \boldsymbol{x}^{(k)}$$

is itself differentiable. Consequently, unrolling K times yields a chain of such transformations, implying the final reconstruction $\boldsymbol{x}^{(K)}$ is a differentiable function of the initial estimate $\boldsymbol{x}^{(0)}$. Formally,

$$f : \boldsymbol{x}^{(0)} \;\mapsto\; \boldsymbol{x}^{(K)} \;=\; \Big(\underbrace{DC \circ \mathcal{F}^{(k)} \circ \cdots \circ DC \circ \mathcal{F}^{(1)}}_{K \text{ times}}\Big)(\boldsymbol{x}^{(0)}).$$

By the chain rule in multivariate calculus, this implies that f has a well-defined Jacobian $\boldsymbol{J}_f \in \mathbb{C}^{n \times n}$ at each input $\boldsymbol{x}^{(0)}$.

C. Verification of Random Vector Properties

Here, we first demonstrate that the introduced random vectors—complex Gaussian and Rademacher—satisfies the conditions stipulated in Theorem 3.1: zero mean and unit covariance.

C.1. Complex Gaussian Vectors

Definition C.1. A **Complex Gaussian vector** $\boldsymbol{v} \in \mathbb{C}^n$ is defined such that each element v_j is independently sampled from a complex normal distribution with zero mean and unit variance, denoted as $v_j \sim \mathcal{CN}(0, 1)$. This implies:

$$v_j = a_j + ib_j,$$

where a_j and b_j are independent real-valued Gaussian random variables with $a_j, b_j \sim \mathcal{N}(0, 0.5)$.

Algorithm 1 Naive Per-Voxel Variance Calculation (Row-by-Row Jacobian)

Require:
- Reconstruction function $f : \mathbb{C}^n \to \mathbb{C}^n$ with Jacobian $\boldsymbol{J}_f \in \mathbb{C}^{n\times n}$,
- Imaging operator $\boldsymbol{A} \in \mathbb{C}^{m\times n}$,
- k-space sample covariance matrix $\boldsymbol{\Sigma}_k \in \mathbb{C}^{m\times m}$.

Ensure: Voxel-wise variance map $\left\{\mathrm{Var}(\boldsymbol{x}_i)\right\}_{i=1}^{n}$.

Compute the Cholesky factor of $\boldsymbol{\Sigma}_k$:

$$\boldsymbol{\Sigma}_k = \boldsymbol{\sigma}_k \, \boldsymbol{\sigma}_k^H.$$

Initialize an empty array for variances:

$$\mathrm{Var}(\boldsymbol{x}) = \mathbf{0} \in \mathbb{R}^n.$$

for each voxel $i = 1, \ldots, n$ **do**

Create one-hot vector $\mathbf{e}_i \in \mathbb{R}^n$ with 1 at index i, 0 otherwise.

Backpropagate $\mathbf{e}_i$ through f to get row i of $\boldsymbol{J}_f$:

$$\nabla^\top f_i \;\leftarrow\; \left(\boldsymbol{J}_f^H \, \mathbf{e}_i\right)^\top.$$

Form

$$\boldsymbol{l}_i \;=\; (\nabla^\top f_i)\, \boldsymbol{A}^H \, \boldsymbol{\sigma}_k.$$

{Imaging operator and noise correlations included.}

Compute variance:

$$\mathrm{Var}(\boldsymbol{x}_i) \;=\; \|\boldsymbol{l}_i\|_2^2.$$

Store in $\mathrm{Var}(\boldsymbol{x})_i \leftarrow \mathrm{Var}(\boldsymbol{x}_i)$.

end for

C.1.1. Zero Mean

The mean of each element v_j is calculated as:

$$\mathbb{E}[v_j] = \mathbb{E}[a_j] + i\mathbb{E}[b_j] = 0 + i \cdot 0 = 0$$

Therefore,

$$\mathbb{E}[\boldsymbol{v}] = \mathbf{0}$$

C.1.2. Unit Covariance

Since a_j and b_j are independent and each has a variance of 0.5, the covariance between v_j and v_k is:

$$\mathbb{E}[v_j v_k^H] = \mathbb{E}[(a_j + ib_j)(a_k - ib_k)] = \mathbb{E}[a_j a_k] + \mathbb{E}[b_j b_k]$$

Given that a_j and b_j are independent across different indices:

$$\mathbb{E}[v_j v_k^H] = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{if } j \neq k. \end{cases}$$

Thus,

$$\mathbb{E}[\boldsymbol{v}\boldsymbol{v}^H] = \boldsymbol{I}$$

C.2. Complex Rademacher Vectors

Definition C.2. A **Complex Rademacher vector** $\boldsymbol{v} \in \mathbb{C}^n$ is defined such that each element v_j is independently sampled from the set $\{e^{j\theta} \mid \theta \sim \text{Uniform}[0, 2\pi)\}$. In other words, each v_j is of the form:

$$v_j = e^{j\theta_j},$$

where θ_j is uniformly distributed over the interval $[0, 2\pi)$.

C.2.1. ZERO MEAN

Since θ_j is uniformly distributed over $[0, 2\pi)$, the expectation of each v_j is:

$$\mathbb{E}[v_j] = \mathbb{E}[e^{j\theta_j}] = \frac{1}{2\pi}\int_0^{2\pi} e^{j\theta} d\theta = 0$$

Therefore,

$$\mathbb{E}[\boldsymbol{v}] = \mathbf{0}$$

C.2.2. UNIT COVARIANCE

For $j = k$,

$$\mathbb{E}[v_j v_j^H] = \mathbb{E}[e^{j\theta_j} e^{-j\theta_j}] = \mathbb{E}[1] = 1$$

For $j \neq k$,

$$\mathbb{E}[v_j v_k^H] = \mathbb{E}[e^{j\theta_j} e^{-j\theta_k}] = \mathbb{E}[e^{j\theta_j}] \cdot \mathbb{E}[e^{-j\theta_k}] = 0 \cdot 0 = 0$$

This separation of expectations is valid due to the independence of θ_j and θ_k. Hence,

$$\mathbb{E}[\boldsymbol{v}\boldsymbol{v}^H] = \boldsymbol{I}$$

D. Variance of Complex Gaussian vs. Complex Random-Phase

Let $\boldsymbol{\Sigma} \in \mathbb{C}^{n\times n}$ be a Hermitian positive semi-definite matrix (e.g., a covariance matrix), and let $\boldsymbol{v} \in \mathbb{C}^n$ be a random vector satisfying

$$\mathbb{E}[\boldsymbol{v}] \;=\; \mathbf{0}, \quad \mathbb{E}\big[\boldsymbol{v}\,\boldsymbol{v}^H\big] \;=\; \boldsymbol{I}.$$

We define the following unbiased estimator for the diagonal entry Σ_{ii}:

$$Y_i \;=\; \big(\boldsymbol{\Sigma}\,\boldsymbol{v}\big)_i \,\big(v_i\big)^*, \quad \Longrightarrow \quad \mathbb{E}[Y_i] \;=\; \Sigma_{ii}.$$

Below, we derive the variance of Y_i under two complex random-vector distributions: *(i) complex Gaussian* and *(ii) complex random-phase (Rademacher)*. In both cases, the key difference lies in the *fourth moment* of the entries.

D.1. General Expansion of $|Y_i|^2$

First, we write out

$$Y_i \;=\; \sum_{j=1}^{n} \Sigma_{ij}\, v_j \;\times\; \big(v_i\big)^*.$$

Hence,

$$|Y_i|^2 \;=\; \Big(\sum_{j=1}^{n} \Sigma_{ij}\, v_j\Big)\,\big(v_i\big)^* \;\times\; \Big(\sum_{k=1}^{n} \Sigma_{ik}\, v_k\Big)^*\,\big(v_i\big)^{**}.$$

Since $\boldsymbol{\Sigma}$ is Hermitian, $\Sigma_{ik}^* = \Sigma_{ki}$, and $(v_i)^{**} = v_i$. Expanding yields

$$|Y_i|^2 \;=\; \Big(\sum_{j=1}^{n} \Sigma_{ij}\, v_j\Big)\,\Big(\sum_{k=1}^{n} \Sigma_{ik}\, v_k\Big)^* \;\times\; |v_i|^2.$$

Taking expectation,

$$\mathbb{E}\big[|Y_i|^2\big] \;=\; \sum_{j=1}^{n}\sum_{k=1}^{n} \Sigma_{ij}\,\Sigma_{ik}^*\;\mathbb{E}\big[v_j\, v_k^*\,|v_i|^2\big].$$

Because $\mathbb{E}[\boldsymbol{v}\,\boldsymbol{v}^H] = \boldsymbol{I}$, we know $\mathbb{E}[v_j\, v_k^*] = \delta_{jk}$. For $j \neq i$, $k \neq i$, the entries v_j and v_i are uncorrelated in magnitude, so $\mathbb{E}[\,|v_i|^2\, v_j\, v_k^*\big] = \mathbb{E}[|v_i|^2]\,\delta_{jk} = 1 \cdot \delta_{jk}$. When $j = i$ or $k = i$, we retain the *fourth moment* term $\mathbb{E}[\,|v_i|^4\,]$. Thus, generically,

$$\mathbb{E}\big[|Y_i|^2\big] \;=\; \sum_{j\neq i} |\Sigma_{ij}|^2\;\mathbb{E}\big[|v_j|^2\,|v_i|^2\big] \;+\; |\Sigma_{ii}|^2\,\mathbb{E}\big[|v_i|^4\big].$$

Given $\mathbb{E}\big[|v_j|^2\big] = 1$ for all j, we ultimately get

$$\mathbb{E}\big[|Y_i|^2\big] \;=\; \sum_{j\neq i}|\Sigma_{ij}|^2 \;+\; |\Sigma_{ii}|^2\,\mathbb{E}\big[|v_i|^4\big].$$

Subtracting $|\Sigma_{ii}|^2$ (which is $\big|\mathbb{E}[Y_i]\big|^2$) then gives the variance:

$$\mathrm{Var}[Y_i] \;=\; \Big[\sum_{j\neq i}|\Sigma_{ij}|^2 \;+\; |\Sigma_{ii}|^2\,\mathbb{E}\big[|v_i|^4\big]\Big] \;-\; |\Sigma_{ii}|^2.$$

Thus the crux is: $\mathbb{E}[|v_i|^4]$.

D.2. Complex Gaussian Case

Distribution. Let each v_i be drawn from a circularly symmetric Gaussian ($\mathcal{CN}(0,1)$), implying

$$\mathbb{E}\big[|v_i|^2\big] = 1, \quad \mathbb{E}\big[|v_i|^4\big] = 2.$$

(That is, if $v_i = \frac{1}{\sqrt{2}}(x_i + j\,y_i)$ with $x_i, y_i \sim \mathcal{N}(0,1)$ i.i.d.)

Variance Calculation. Substitute $\mathbb{E}[|v_i|^4] = 2$ into the generic formula:

$$\mathbb{E}\big[|Y_i|^2\big] \;=\; \sum_{j\neq i}|\Sigma_{ij}|^2 \;+\; 2\,|\Sigma_{ii}|^2.$$

Therefore,

$$\mathrm{Var}[Y_i] \;=\; \sum_{j\neq i}|\Sigma_{ij}|^2 \;+\; 2\,|\Sigma_{ii}|^2 \;-\; |\Sigma_{ii}|^2 \;=\; \sum_{j=1}^{n}|\Sigma_{ij}|^2.$$

Hence, for *complex Gaussian*:

$$\mathrm{Var}[Y_i] \;=\; \sum_{j=1}^{n}\big|\Sigma_{ij}\big|^2.$$

D.3. Complex Random-Phase (Rademacher) Case

Distribution. Now let each v_i have *unit magnitude* and uniformly random phase,

$$|v_i| = 1, \quad v_i \;=\; e^{j\,\theta_i}, \quad \theta_i \sim \mathrm{Uniform}[0, 2\pi].$$

Then

$$\mathbb{E}\big[|v_i|^4\big] \;=\; 1.$$

Variance Calculation. Substituting $\mathbb{E}[|v_i|^4] = 1$ into our generic formula:

$$\mathbb{E}\big[|Y_i|^2\big] \;=\; \sum_{j\neq i}|\Sigma_{ij}|^2 \;+\; 1\cdot|\Sigma_{ii}|^2 \;=\; \sum_{j=1}^{n}|\Sigma_{ij}|^2.$$

Subtracting $|\Sigma_{ii}|^2$ from this, we get

$$\mathrm{Var}[Y_i] \;=\; \sum_{j=1}^{n}|\Sigma_{ij}|^2 \;-\; |\Sigma_{ii}|^2.$$

Thus, for *complex random-phase* (or "complex Rademacher"):

$$\mathrm{Var}[Y_i] \;=\; \sum_{j=1}^{n}|\Sigma_{ij}|^2 \;-\; |\Sigma_{ii}|^2.$$

E. Bounds on the Estimation (sketching) Error

Theorem E.1 (Bounds on the Sketching Error)**.** *Let $\mathbf{\Sigma}_x \in \mathbb{C}^{n\times n}$ be a Hermitian positive semidefinite matrix and define its off-diagonal component by*

$$\overline{\mathbf{\Sigma}}_x = \mathbf{\Sigma}_x - \mathrm{diag}(\mathbf{\Sigma}_x).$$

Let $r_m(\mathbf{\Sigma}_x)$ be the diagonal estimator obtained by averaging m independent complex random-phase (Rademacher) sketches. Then, for any $\delta \in (0, 1]$, with probability at least $1 - \delta$,

$$\|r_m(\mathbf{\Sigma}_x) - \mathrm{diag}(\mathbf{\Sigma}_x)\|_2 \le c\sqrt{\frac{\log(2/\delta)}{m}}\,\|\overline{\mathbf{\Sigma}}_x\|_F,$$

where $c > 0$ is an absolute constant independent of $\mathbf{\Sigma}_x$ and m.

Proof. Define the error vector

$$e = m\Big(r_m(\mathbf{\Sigma}_x) - \mathrm{diag}(\mathbf{\Sigma}_x)\Big).$$

By construction, e is the sum of m i.i.d. mean-zero random vectors:

$$e = \sum_{z=1}^{m} e_z,$$

where each e_z arises from a single sketch of our estimator for $\mathbf{\Sigma}_x$. In our case, by design the expected squared norm of the error in a single sketch is given by

$$\mathbb{E}\|e_z\|_2^2 = \|\overline{\mathbf{\Sigma}}_x\|_F^2.$$

The key observation is that the variance of the scalar estimator for the ith diagonal entry is (From Appendix D.3):

$$\mathrm{Var}[Y_i] = \sum_{j=1}^{n} |\Sigma_{x,ij}|^2 - |\Sigma_{x,ii}|^2,$$

so that the contribution of the diagonal is subtracted, yielding a tighter bound.

We now follow a symmetrization and scalar-comparison argument (Vershynin, 2018)):

1. **Symmetrization.** Let e' be an independent copy of e. Since $f(x) = \|x\|_2^2$ is convex and $\mathbb{E}[e] = 0$, by Jensen's inequality we have

$$\mathbb{E}\|e\|_2^{2k} \le \mathbb{E}\|e - e'\|_2^{2k}. \tag{15}$$

2. Scalar Comparison. For each sketch index z, let $\mathbf{e}_{z,1}$ and $\mathbf{e}_{z,2}$ be independent copies of the single-sketch error $\mathbf{e}_z \in \mathbb{C}^n$, and define

$$\mathbf{w}_z \;=\; \mathbf{e}_{z,1} - \mathbf{e}_{z,2}.$$

Although $\mathbf{w}_z$ may have complex entries (due to complex random-phase vectors in the estimation process), its squared norm $\|\mathbf{w}_z\|_2^2$ is a *real, nonnegative* scalar.

Next, introduce an independent *real* Rademacher random variable $r_z \in \{-1, +1\}$ for each z, and define

$$W_z \;=\; r_z\,\|\mathbf{w}_z\|_2^2.$$

Because $\mathbf{w}_z$ has a symmetric (mean-zero) distribution in $\mathbb{C}^n$, one can show that $\mathbb{E}[W_z] = 0$. Moreover, for any integer $k \ge 1$, since $r_z^{2k} = 1$, we obtain

$$\mathbb{E}[W_z^{2k}] \;=\; \mathbb{E}\Big[\big(r_z\,\|\mathbf{w}_z\|_2^2\big)^{2k}\Big] \;=\; \mathbb{E}\big[\|\mathbf{w}_z\|_2^{4k}\big].$$

Let

$$E \;=\; \sum_{z=1}^{m} W_z \;=\; \sum_{z=1}^{m}\Big(r_z\,\|\mathbf{w}_z\|_2^2\Big).$$

Algorithm 2 Noise Calculation in DL-based MRI reconstruction via Sketching the Network Jacobian

Require:
- Imaging operator $\boldsymbol{A} \in \mathbb{C}^{m\times n}$,
- k-space sample covariance matrix $\boldsymbol{\Sigma}_k \in \mathbb{C}^{m\times m}$
- Reconstruction network $f : \mathbb{C}^n \to \mathbb{C}^n$ with Jacobian $\boldsymbol{J}_f$,
- S, size of random matrix $\boldsymbol{V}_S$.

Ensure: Approximate $\widehat{\operatorname{diag}\left(\boldsymbol{\Sigma}_{\boldsymbol{x}}\right)} \approx \operatorname{diag}\left(\boldsymbol{\Sigma}_{\boldsymbol{x}}\right)$, where $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{J}_f\, \boldsymbol{A}^H\, \boldsymbol{\Sigma}_k\, \boldsymbol{A}\, \boldsymbol{J}_f^H$

Compute Cholesky factorization: $\boldsymbol{\Sigma}_k = \boldsymbol{\sigma}_k \boldsymbol{\sigma}_k^H$

$\boldsymbol{V}_S \in \mathbb{C}^{m\times S}$ with columns $\{\boldsymbol{v}_j\}$ sampled from a *complex random-phase* distribution:

$$v_i^{(j)} \;=\; e^{\,j\,\theta_i^{(j)}}, \quad \theta_i^{(j)} \sim \text{Uniform}[0, 2\pi].$$

$\boldsymbol{W}_S \leftarrow \boldsymbol{\sigma}_k\, \boldsymbol{V}_S \quad (\in \mathbb{C}^{m\times S})$

$\widetilde{\boldsymbol{W}}_S \leftarrow \boldsymbol{A}^H\, \boldsymbol{W}_S \quad (\in \mathbb{C}^{n\times S})$

$\boldsymbol{U}_S \leftarrow \boldsymbol{J}_f\big(\widetilde{\boldsymbol{W}}_S\big)$

$\boldsymbol{V}_{\text{samples}} \leftarrow \boldsymbol{U}_S \odot \boldsymbol{U}_S^H \quad (\in \mathbb{R}^{n\times S})$

$\widehat{\operatorname{diag}\left(\boldsymbol{\Sigma}_{\boldsymbol{x}}\right)} \;\leftarrow\; \frac{1}{S}\, \boldsymbol{V}_{\text{samples}}\, \mathbf{1}_S \quad (\in \mathbb{R}^n)$

A careful term-by-term expansion (exploiting the symmetry of each $\mathbf{w}_z \mapsto -\mathbf{w}_z$) shows that

$$\mathbb{E}\big[\|\mathbf{e}\|_2^{2k}\big] \;\;\le\;\; \mathbb{E}\big[\big|E\big|^{2k}\big].$$

Hence, we reduce bounding $\|\mathbf{e}\|_2$ (a vector norm in $\mathbb{C}^n$) to analyzing E (a sum of real scalars). This paves the way for standard sub-Gaussian/sub-Exponential tail bounds to be applied in a straightforward manner.

3. **Concentration.** Standard hypercontractivity and concentration results (Roosta-Khorasani & Ascher, 2013)) imply that E is sub-Gaussian with parameter

$$K = C\sqrt{m}\,\|\overline{\boldsymbol{\Sigma}}_x\|_F,$$

for an absolute constant $C > 0$. Consequently, E^2 is sub-exponential with parameter proportional to $m\,\|\overline{\boldsymbol{\Sigma}}_x\|_F^2$. In particular, applying Proposition 2.7.1 from (Vershynin, 2018), we obtain that for any $t > 0$,

$$\Pr\left(\frac{\|e\|_2^2}{m^2} \ge t\right) \le 2\exp\left(-\frac{t}{c_0\,\|\overline{\boldsymbol{\Sigma}}_x\|_F^2}\right),$$

for some constant $c_0 > 0$.

Setting

$$t = c_0\,\|\overline{\boldsymbol{\Sigma}}_x\|_F^2 \log\frac{2}{\delta},$$

we deduce that with probability at least $1-\delta$,

$$\frac{\|e\|_2^2}{m^2} \le c_0\,\|\overline{\boldsymbol{\Sigma}}_x\|_F^2 \log\frac{2}{\delta}.$$

Recalling that $e = m\Big(r_m(\boldsymbol{\Sigma}_x) - \operatorname{diag}(\boldsymbol{\Sigma}_x)\Big)$, it follows that

$$\|r_m(\boldsymbol{\Sigma}_x) - \operatorname{diag}(\boldsymbol{\Sigma}_x)\|_2 \le \sqrt{c_0}\,\sqrt{\frac{\log(2/\delta)}{m}}\,\|\overline{\boldsymbol{\Sigma}}_x\|_F.$$

Renaming $\sqrt{c_0}$ as c completes the proof. □

F. Sketching Matrix

Figure 4 varies S, size of the sketching matrix $\boldsymbol{V}_S$, and measures the resulting NRMSE of our variance estimator. As S increases from 100 to 1,900, there is a subtle decrease in error. This trend reflects the improved accuracy of the diagonal estimator when more random vectors probe the Jacobian structure. However, beyond a certain point (approximately $S = 1,000$–$1,200$ in our experiments), the NRMSE decreases more gradually, indicating diminishing returns in accuracy. Thus, while a large S can yield more precise variance maps, it also incurs higher computational cost; our choice of $S = 1,000$ would yield a favorable balance between accuracy and runtime for most practical settings.

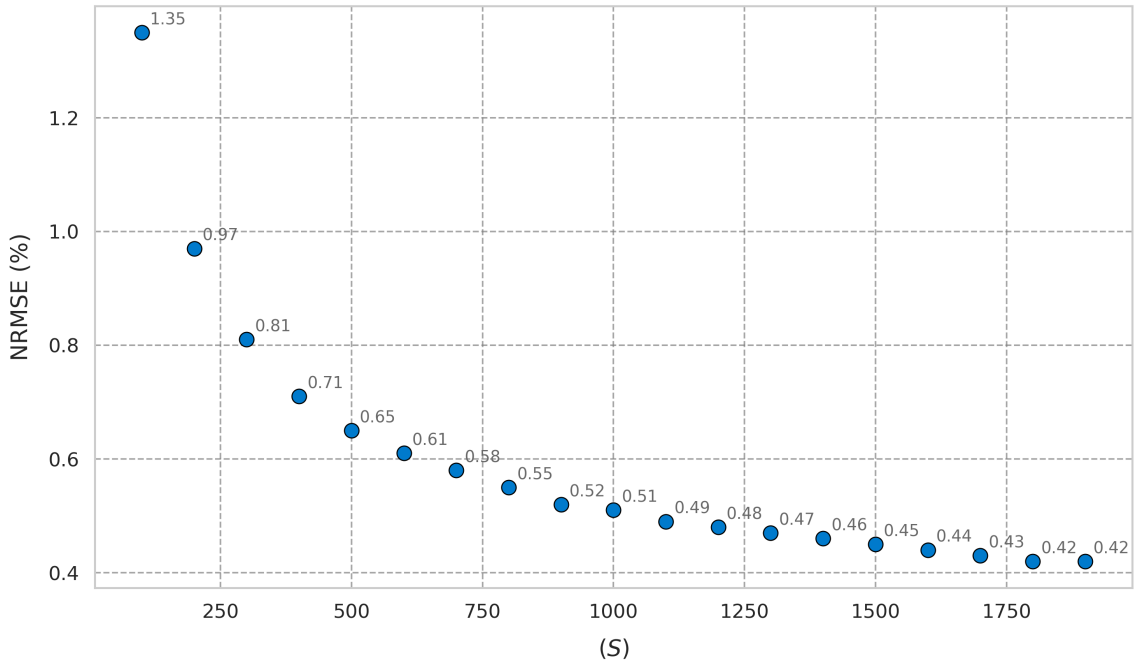

Figure 4. NRMSE of the proposed method over test slices vs. size of the sketching matrix (S) (E2E-VarNet on brain data at $R = 8, \alpha = 1$). As the size S of the sketching matrix $\boldsymbol{V}_S$ grows from 100 to 1,900, the Normalized Root-Mean-Square Error (NRMSE) of the variance estimator decreases. While additional probing vectors generally improve accuracy by capturing more of the Jacobian's structure, the marginal benefit tapers off beyond about 1,000–1,200 vectors, suggesting a practical trade-off between improved precision and increased computation time.

G. Empirical Noise Calculation

MC Trials. For each experiment (reconstruction method, noise level, acceleration rate, dataset) and slice, we simulated $N = 3,000$ MC trials. Specifically, we draw the noise vector $\boldsymbol{n}$ from $\boldsymbol{\Sigma}_k$ for each trial, add it to the measured k-space, then reconstruct the image. Across these 3,000 reconstructions, we compute voxelwise sample variance to form an *empirical* variance map, which serves as the "gold standard" reference.

Convergence. Although we typically fix 3,000 trials to ensure a stable reference variance, Figure 5 illustrates the effect of varying the trial count from 100 to 5,000. For smaller trial counts, the empirical variance estimate shows notably larger NRMSE relative to a high-sample baseline (e.g., 10,000 trials), but this error diminishes as we approach 2,500–3,000 trials, thereafter offering only gradual improvements. Consequently, while increasing the number of MC samples monotonically improves estimation accuracy, it quickly becomes computationally prohibitive to reconstruct each perturbed k-space many thousands of times.

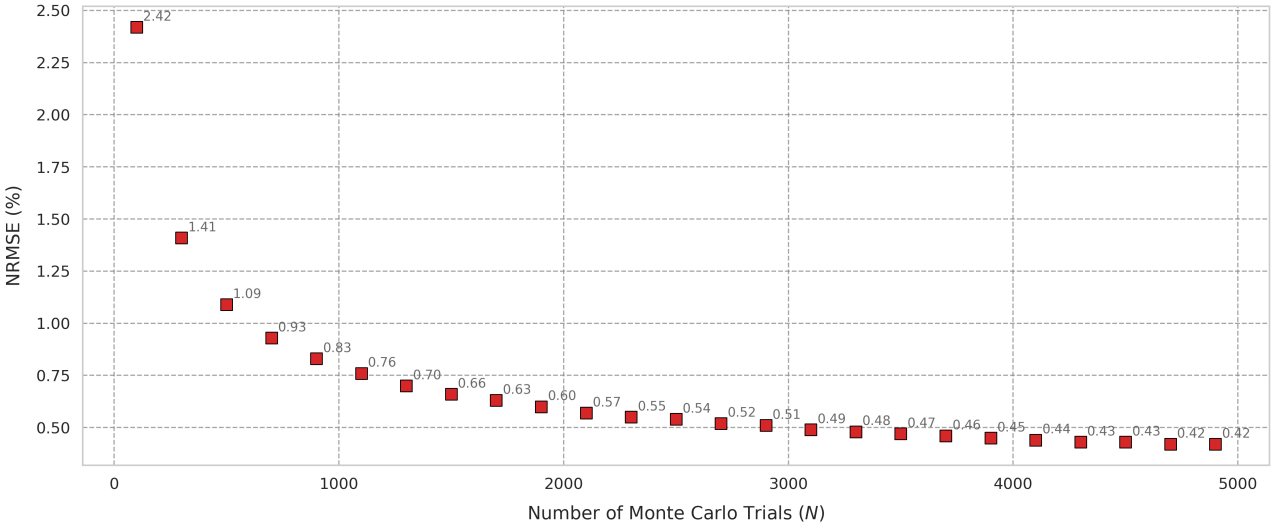

Figure 5. Mean NRMSE of empirical baseline over test slices vs. the number of MC Trials (N) (E2E-VarNet on brain data at $R = 8, \alpha = 1$). We vary the total number of MC realizations from 100 up to 5,000 and compare the resulting variance maps to a high-sample reference (e.g., 10,000 trials). As more trials are included, the empirical variance estimation converges around $N = 3000$, shown here by the decreasing NRMSE.

H. Reconstruction Methods

Here we provide detailed descriptions of the six deep MRI reconstruction algorithms evaluated in our study. Each algorithm falls under a distinct combination of *learning paradigm* and *architecture design*, ranging from supervised to self-supervised, and from fully data-driven to physics-driven unrolled models. All neural architectures incorporate complex-valued inputs via stacking real and imaginary parts, resulting in 2 input and output channels. Note that hyperparameters for each method were tuned on the validation dataset to ensure optimal performance. Unless otherwise stated, all models were trained via Adam optimizer (Kingma & Ba, 2017), with $\beta_1 = 0.965, \beta_2 = 0.99$. Here, we provide detailed descriptions for each method. We used PyTorch for modeling (Paszke et al., 2019), and NVIDIA NVIDIA RTX A6000 for GPU acceleration.

H.1. Supervised, Physics-Driven Unrolled

H.1.1. E2E-VarNet (Sriram et al., 2020a)

E2E-VarNet is a multi-coil unrolled reconstruction method that alternates between learned regularization blocks and data-consistency (DC) steps. Each iteration applies a trainable neural network (often a ResNet or similar CNN) to refine the current image estimate, followed by a DC step enforcing agreement with the measured multi-coil *k*-space data.

We configure E2E-VarNet with 4 unrolled steps, each containing 2 ResNet blocks (256 channels, kernel size 3). An ℓ_1-norm reconstruction loss is used, and the network is optimized with Adam (lr=10^{-4}). Table 3 summarizes additional hyperparameters for knee and brain MRI datasets.

H.1.2. MoDL (Aggarwal et al., 2019)

MoDL (Model-Based Deep Learning) integrates conjugate-gradient (CG) steps for data-consistency with a CNN denoiser block for learned regularization. Each unrolled iteration iteratively solves a least-squares subproblem via CG and then applies a trainable ResNet denoiser.

We set MoDL to 4 unrolled iterations, each containing 10 internal CG steps. A ResNet (256 channels, 2 blocks per iteration) serves as the learned regularizer. Table 4 details these hyperparameters, including ℓ_1 reconstruction loss, Adam (lr=10^{-4}), and separate CG tolerances for knee vs. brain.

Table 3. Key Hyperparameters used for training E2E-VarNet on knee and brain datasets.

Hyperparameter	**Knee**	**Brain**
Architecture		
Unrolled Steps	4	4
Block Architecture	ResNet	ResNet
ResNet Channels	256	256
# ResNet Blocks / Step	2	2
Kernel Size	3	3
Training		
Loss Function	ℓ_1	ℓ_1
Batch Size	4	4
Grad. Accum. Steps	4	4
Base Learning Rate	10^{-4}	10^{-4}
Weight Decay	10^{-4}	10^{-4}
Max Iterations	120,000	50,000

Table 4. Key Hyperparameters used for training MoDL on knee and brain datasets.

Hyperparameter	**Knee**	**Brain**
Architecture		
Unrolled Steps	4	4
DC Solver	CG (10 iterations, $\epsilon = 10^{-5}$)	CG (10 iterations, $\epsilon = 10^{-4}$)
Block Architecture	ResNet	ResNet
ResNet Channels	256	256
# ResNet Blocks / Step	2	2
Kernel Size	3	3
Training		
Loss Function	ℓ_1	ℓ_1
Batch Size	2	2
Grad. Accum. Steps	8	4
Base Learning Rate	10^{-4}	10^{-4}
Weight Decay	10^{-4}	10^{-4}
Max Iterations	80,000	80,000

H.2. Supervised, Data-Driven

H.2.1. U-NET (RONNEBERGER ET AL., 2015)

Unlike unrolled methods that tightly integrate the MRI forward model with iterative DC, **U-Net** is a fully data-driven, "one-shot" approach, mapping the zero-filled (or coil-combined) image directly to the reconstructed output. This yields faster inference, but it lacks physics-based constraints.

We use a standard U-Net with 4 levels of down/up-sampling, 32 base channels, and blocks of {`conv, relu, conv, relu, batchnorm, dropout`}. An ℓ_1 reconstruction loss and Adam (lr=10^{-3}) train the network. Table 5 summarizes key parameters for knee and brain datasets.

H.3. Semi-Supervised, Physics-Driven Unrolled

H.3.1. N2R (NOISE2RECON) (DESAI ET AL., 2023)

N2R leverages both a limited set of fully sampled ground-truth data and a larger set of undersampled data to train an unrolled network. Each iteration alternates between DC blocks and learned ResNet-based regularization blocks, enforcing consistency only on sampled *k*-space points. Table 6 outlines key hyperparameters, including the *k*-space ℓ_1 loss, 4 unrolled steps, 256-channel ResNet blocks, and consistency noise levels ([0.2, 0.5]) .

Table 5. Key Hyperparameters used for training U-Net on knee and brain datasets.

Hyperparameter	**Knee**	**Brain**
Architecture		
Base Channels	32	32
Depth (Pool Layers)	4	4
Block Sequence	{conv, relu, conv, relu, bn, dropout}	{conv, relu, conv, relu, bn, dropout}
Training		
Loss Function	ℓ_1	ℓ_1
Batch Size	16	12
Grad. Accum. Steps	1	2
Base Learning Rate	10^{-3}	10^{-3}
Weight Decay	10^{-4}	10^{-4}
Training Length	200 epochs	80,000 iterations

Table 6. Key hyperparameters used for training N2R on knee and brain datasets.

Hyperparameter	**Knee**	**Brain**
Architecture		
Total Training Scans	14 (13 undersamp.)	54 (50 undersamp.)
Acceleration Factor	8	8
Unrolled Steps	4	4
Block Architecture	ResNet (256 ch)	ResNet (256 ch)
# ResNet Blocks / Step	2	2
Kernel Size	3	3
Semi-Supervision		
Consistency noise std σ Range	[0.2, 0.5]	[0.2, 0.5]
Training		
Loss Function	k-space ℓ_1	k-space ℓ_1
Batch Size	2	1
Grad. Accum. Steps	8	4
Base Learning Rate	10^{-4}	10^{-4}
Weight Decay	10^{-4}	10^{-4}
Max Iterations	80,000	80,000

H.3.2. VORTEX (DESAI ET AL., 2022A)

VORTEX is another semi-supervised unrolled method, leveraging partial fully-sampled references alongside motion/noise augmentations. Each iteration incorporates DC steps and 256-channel ResNet blocks. Random motion/noise transformations align with MR physics, enforcing consistent reconstructions under both fully and partially sampled data.

Like N2R, VORTEX employs 4 unrolled steps (2 ResNet blocks/step, 256 channels, kernel size=3). Loss is a mixture of ℓ_1 and ℓ_2 in the k-space domain. Table 7 summarizes these hyperparameters, showing how motion/noise augmentations are tuned slightly differently between knee vs. brain data.

H.3.3. SSDU (YAMAN ET AL., 2020)

Self-Supervised Learning via Data Undersampling (SSDU) obviates the need for fully sampled targets by splitting acquired k-space data into "label" and "mask" sets. The "label" portion defines the reconstruction loss, while the "mask" portion is used within DC steps during the unrolled network's forward pass. Mask ratio $\rho = 0.2$ partitions 20% of acquired samples as "label," with the remaining 80% guiding DC.

SSDU's unrolled network matches E2E-VarNet's structure (4 steps, ResNet blocks at 256 channels). All training references were undersampled, partitioning k-space into label vs. mask sets. Table 8 outlines the key hyperparameters, highlighting the uniform mask, ℓ_1 k-space loss, and typical Adam optimization (lr=10^{-4}).

Table 7. Key hyperparameters used for training VORTEX on knee and brain datasets.

Hyperparameter	**Knee**	**Brain**
Architecture		
Acceleration Factor	8	8
Total Training Scans	14 (13 undersamp.)	54 (50 undersamp.)
Unrolled Steps	4	4
Block Architecture	ResNet (256 ch)	ResNet (256 ch)
# ResNet Blocks / Step	2	2
Kernel Size	3	3
VORTEX Motion/Noise Augmentations		
Motion Severity (α) Range	[0.2, 0.5]	[0.1, 0.2]
Noise STD (σ) Range	[0.2, 0.5]	[0.1, 0.2]
Training		
Loss Function	k-space ℓ_1/ℓ_2	k-space ℓ_1/ℓ_2
Batch Size	1	1
Grad. Accum. Steps	4	4
Base Learning Rate	10^{-4}	10^{-4}
Weight Decay	10^{-4}	10^{-4}
Max Iterations	80,000	80,000

Table 8. Key hyperparameters used for training SSDU on knee and brain datasets.

Hyperparameter	**Knee**	**Brain**
Architecture		
Unrolled Steps	4	4
Block Architecture	ResNet (256 ch)	ResNet (256 ch)
# ResNet Blocks / Step	2	2
Kernel Size	3	3
Self-supervision & Masking Setup		
Mask Type	uniform	uniform
ρ (Label Fraction)	0.2	0.2
Training		
Loss Function	k-space ℓ_1	k-space ℓ_1
Batch Size	2	1
Grad. Accum. Steps	8	4
Base Learning Rate	10^{-4}	10^{-4}
Weight Decay	10^{-4}	10^{-4}
Max Iterations	80,000	80,000

H.4. Reconstruction Results

All methods were trained and evaluated on *clean*, non-noisy data. Table 9 lists the mean Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) for knee and brain test sets across the six analyzed reconstruction methods. Overall, E2E-VarNet and MoDL achieve the highest PSNR values on knee scans, whereas N2R and VORTEX perform strongly on brain data—particularly evident in SSIM. U-Net and SSDU yield somewhat lower scores, potentially reflecting their more limited regularization strategies (fully data-driven in U-Net, fully self-supervised in SSDU). Notably, the reconstruction performances are comparable to those reported in the original studies or similar benchmarks, indicating that the networks were trained effectively. Overall, all methods surpass typical parallel imaging baselines in fidelity, confirming sufficient training convergence and general image quality for subsequent noise-variance analyses.

I. Computational Efficiency vs. Network Architecture

Table 10 compares *empirical* versus *proposed* variance method computation times for different neural architectures on the knee dataset. In unrolled frameworks (E2E-VarNet, N2R, VORTEX, SSDU), the computation time increases with more "Steps" (i.e., DC/regularization blocks), reflecting the deeper or more recurrent structure of the unrolled pipeline.

Table 9. Quantitative reconstruction performance across methods on knee and brain datasets

Method	Knee		Brain	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
E2E-VarNet	40.52	0.899	36.31	0.895
MoDL	40.60	0.899	36.44	0.847
U-Net	38.81	0.852	36.12	0.849
N2R	39.52	0.892	36.57	0.918
VORTEX	40.47	0.894	35.78	0.914
SSDU	39.88	0.862	35.51	0.885

Table 10. **Computation Times Across Different Neural Architectures.** Each entry lists the average time (in seconds) required to compute variance maps per slice on the knee dataset, comparing *Empirical* (MC) versus our *Proposed* (Jacobian-sketched) noise calculation. Unrolled architectures vary in steps, indicating the number of DC/regularization blocks. MoDL uses Conjugate Gradient (CG) steps within each DC block, and U-Net is a single-shot CNN without explicit unrolled steps.

Architecture	Variant	Empirical time	Proposed time
Unrolled (E2E-VarNet, N2R, VORTEX, SSDU)	**2 Steps**	27.9	0.2
	4 Steps	54.0	1.3
	6 Steps	73.8	3.4
	8 Steps	97.8	12.1
	10 Steps	120.6	20.6
U-Net	(Ronneberger et al., 2015)	9.6	0.4
Unrolled w/CG-DC (MoDL)	**4 Steps**	181.5	36.9

Nevertheless, our proposed method always remains an order of magnitude faster than empirical references. *U-Net* requires fewer parameters (one-shot CNN without DC blocks), so the empirical approach is already relatively fast, yet we still observe a significant speedup with the proposed method. Finally, *MoDL* includes CG iterations within each DC step, driving up the empirical runtime significantly (181.5 s for 4 steps), while our method's runtime increases more modestly (36.9 s).

These results indicate that both the *unrolled depth* (number of repeated modules) and *type of network layers* (e.g. CG-based DC, convolutional blocks, or potential self-attention units) can influence the total variance-computation time in our framework, albeit to a lesser extent than empirical MC. For instance, substituting convolutional layers with self-attention could raise per-iteration overhead but might not drastically inflate the overall Jacobian-sketch cost, since our approach only requires *Jacobian-vector products* rather than a full Jacobian matrix. Ultimately, the proposed sketching method retains its computational advantage across diverse architectures, enabling efficient variance estimation without imposing strict constraints on layer choices or unrolling depth.

J. Practical Comparison with Low-Trial Monte Carlo

A natural question arises regarding whether empirical reference maps with smaller number of MC trials (e.g., 50–100) might suffice to achieve competitive accuracy without incurring the high computational cost typically associated with MC sampling. To explore this, we compared the proposed sketching method (with sketch size S) and MC (with N trials), each varied over a range from 5 to 130, and measured the NRMSE against a high-accuracy reference map computed using $N = 10{,}000$. Experiments were performed on E2E-VarNet at $R = 8, \alpha = 1$. As shown in Figure 9, MC exhibits large estimation errors for $N \lesssim 50$, whereas our method maintains relatively lower errors even at small S. At larger values of $S \approx N$, our approach remains more accurate while also being faster to compute: for instance, at $S = N = 100$, MC requires ~ 1.8 seconds, whereas our method takes only ~ 0.12 seconds. This behavior aligns with our convergence experiments (see G), indicating that MC variance estimates tend to require three orders of magnitude trials (e.g., $N \approx 1{,}000$–$10{,}000$) to reach a reliable accuracy. In contrast, the sketching-based method attains lower error with at very few sample regimes. Hence, even when MC is reduced to 50–100 trials to speed up computation, the resulting noise estimates suffer in accuracy compared to the proposed sketching approach, highlighting a clear advantage in both performance and speed.

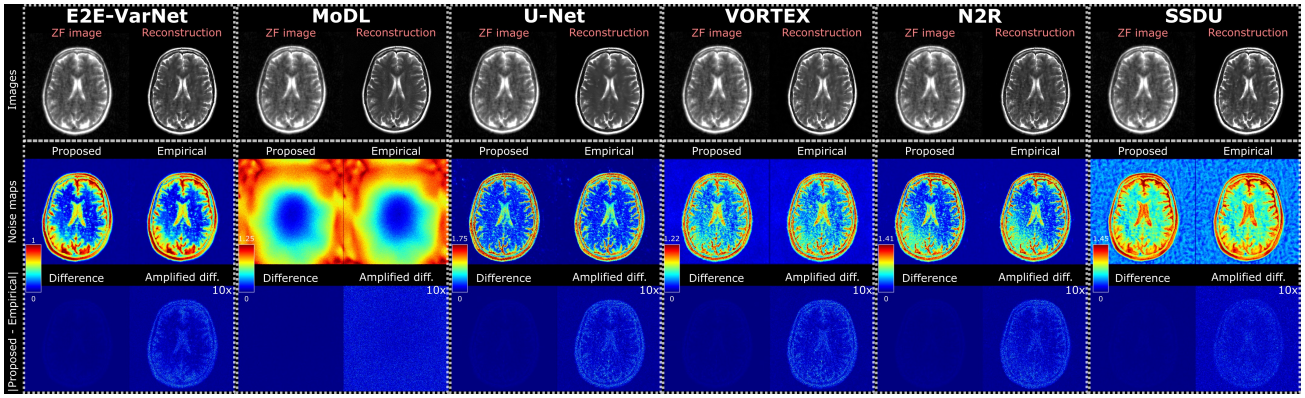

Figure 6. Each column corresponds to a distinct deep reconstruction method (E2E-VarNet, MoDL, U-Net, VORTEX, N2R, SSDU) on brain data at $R = 8$, $\alpha = 1$. In each column: **(top row)** shows the zero-filled (ZF) image and the final reconstructed; **(middle row)** displays variance maps derived by the proposed method and empirical simulations; **(bottom row)** presents the difference map between the proposed and empirical variance estimates, plus an "amplified" difference to further highlight spatial discrepancies. The color bars indicate each map's display window in arbitrary units of variance. Because MoDL integrates an iterative congute-gradient CG-based data consistency step that repeatedly applies the imaging and adjoint operator, the resulting noise map often mirrors the specific coil geometry and undersampling scheme. Regions with poorer coil coverage or unsampled k-space lines accumulate higher residuals, as the iterative solver relies more heavily on the learned prior in those areas. Consequently, MoDL's noise distribution visibly reflects the interplay of coil sensitivities, the mask pattern, and the iterative gradient updates compared to data-dependent Jacobian structure, resulting in the distinct patterns

K. Undersampling Patterns

In this study, we analyzed a range of undersampling schemes to evaluate their impact on noise propagation and variance estimation. These include sampling patterns widely used in musculoskeletal imaging, such as 1D Cartesian uniform and random undersampling, as well as 2D pseudo-random schemes like Poisson Disc and uniform random sampling (Johnson et al., 2023; Desai et al., 2022b) (See Fig. 10. Each pattern introduces distinct aliasing artifacts in the image domain—ranging from coherent aliasing artifacts in uniform Cartesian sampling for parallel imaging to noise-like incoherent artifacts in 2D pseudo-random patterns that are better suited for DL algorithms. Here we investigate these diverse schemes to provide insights into how different undersampling strategies influence the efficacy our noise calculation method (4.4).

L. Additional Results for Robustness to k-space Noise Level

Figure 14 demonstrates tight alignment around the diagonal, exhibiting near-unity R^2 values that confirm a strong match between empirical and theoretical STD at *lower to high noise* ($1\times$–$30\times$). As noise gets excessively larger to the clinically irrelevant regimes ($50\times$–$200\times$), our model *overestimates* the output noise relative to the reference: the best-fit regression line lies above the diagonal, and the scatter broadens. Nevertheless, the correlation remains consistently high, indicating that the model still captures the main trend in noise propagation while modestly overestimating variance in these extreme scenarios.

M. Additional Visualizations of Voxel-wise Noise Estimation

We provide additional visualizations in Figure 12 to offer an improved quantitative perspective on how closely our method aligns with empirical references and where noise tends to concentrate. These visualizations, derived from our voxel-wise noise map estimations, provide further insights. For instance, a histogram of the relative estimation error ((theoretical STD - empirical STD) / empirical STD), as shown in the second panel of Figure 12, reveals the full distribution of voxel-wise deviations, demonstrating that most errors cluster near zero. Furthermore, a scatter plot of the calculated theoretical standard deviation (STD) versus reconstructed voxel intensity, presented in the third panel of Figure 12, shows a moderate positive correlation ($R^2 \approx 0.51$), which is consistent with typical MRI acquisition where noise can scale with signal magnitude. Correspondingly, a scatter plot of the absolute estimation error (the absolute difference between theoretical and empirical

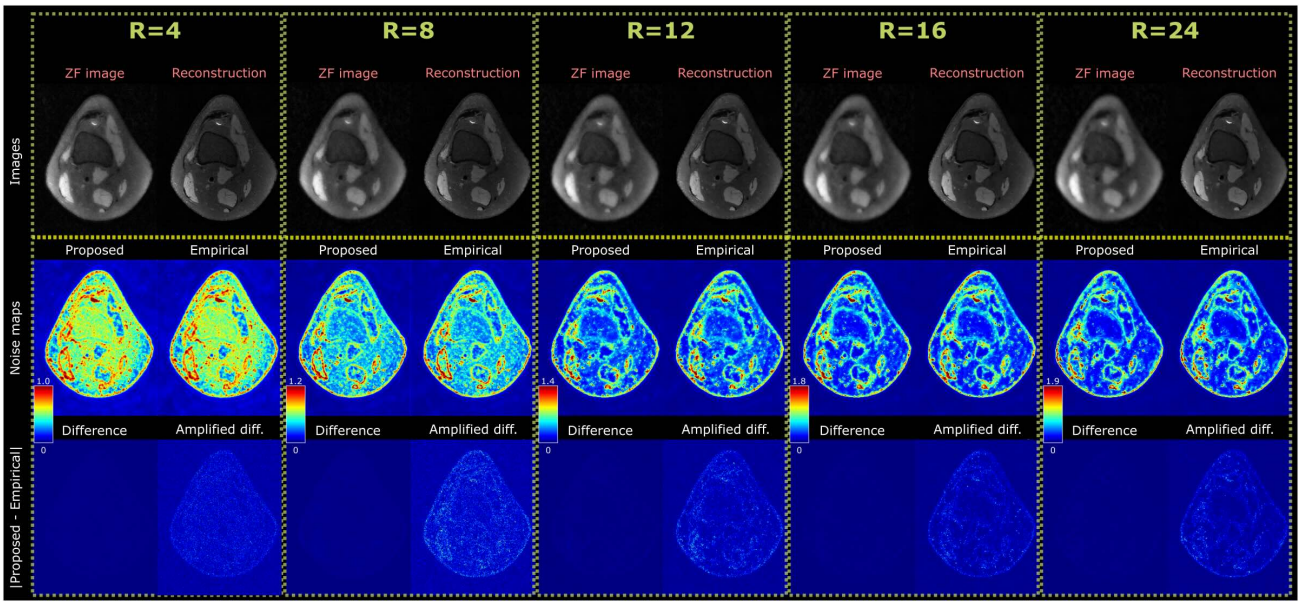

Figure 7. Each column corresponds to a different acceleration factor R (MoDL on knee at $\alpha = 1$). In each column: **(top row)** shows the zero-filled (ZF) image and the final reconstructed; **(middle row)** displays variance maps derived by the proposed method and empirical simulations; **(bottom row)** presents the difference map between the proposed and empirical variance estimates, plus an "amplified" difference to further highlight spatial discrepancies. The color bars indicate each map's display window in arbitrary units of variance.

STD) versus voxel intensity, depicted in the fourth panel of Figure 12, likewise indicates a mild positive correlation ($R^2 \approx 0.16$), suggesting that regions with larger estimation discrepancies can coincide with brighter intensities.

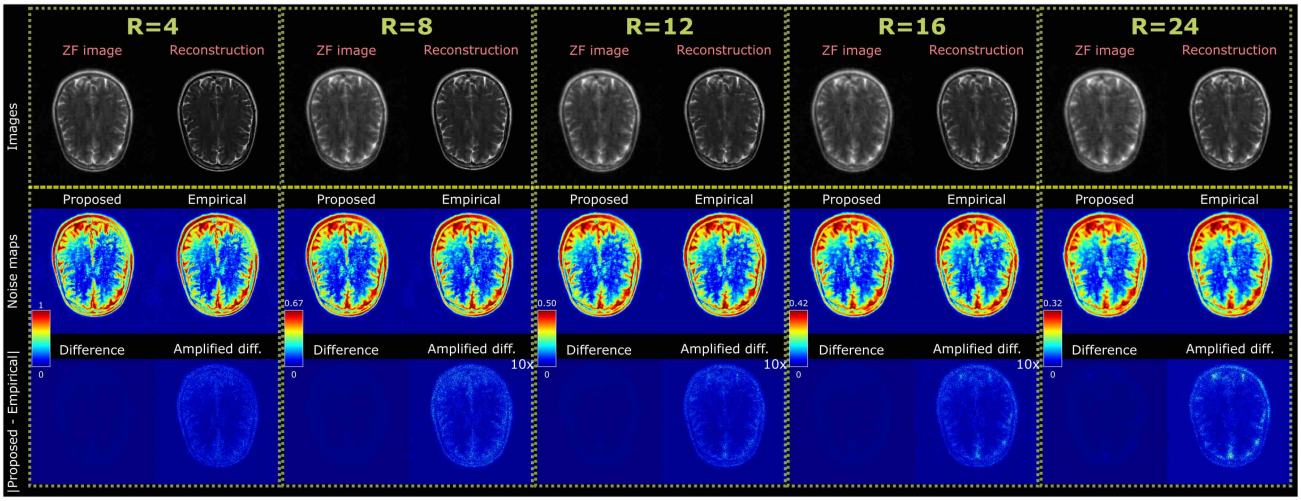

Figure 8. Each column corresponds to a different acceleration factor R (E2E-VarNet on brain at $\alpha = 1$). In each column: **(top row)** shows the zero-filled (ZF) image and the final reconstructed; **(middle row)** displays variance maps derived by the proposed method and empirical simulations; **(bottom row)** presents the difference map between the proposed and empirical variance estimates, plus an "amplified" difference to further highlight spatial discrepancies. The color bars indicate each map's display window in arbitrary units of variance.

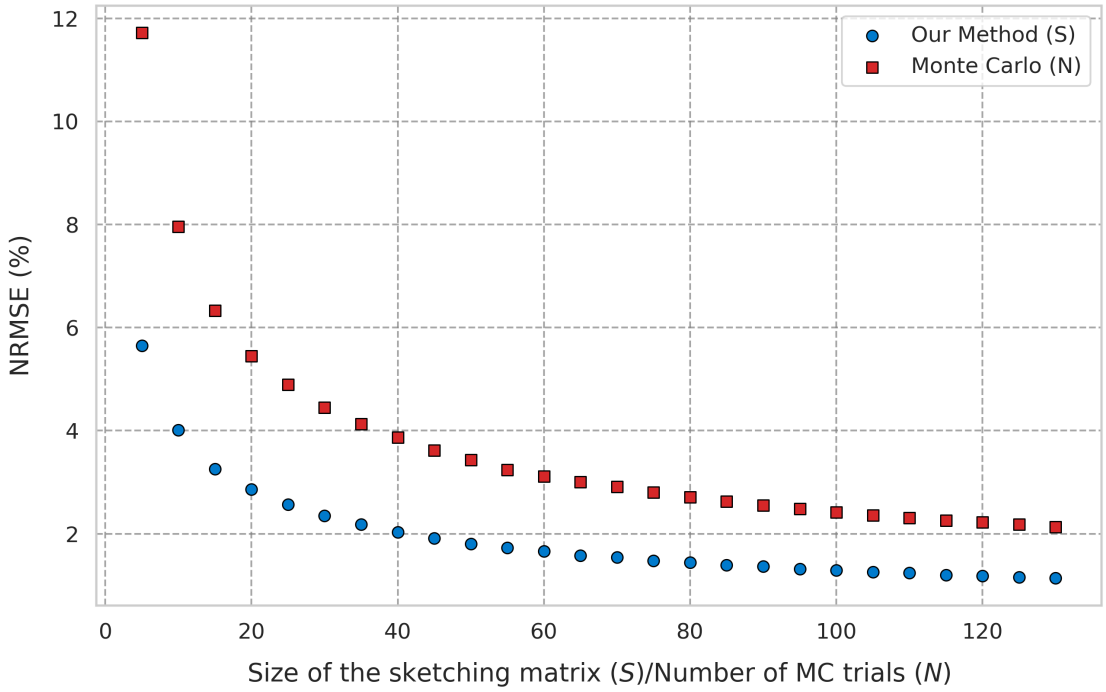

Figure 9. Comparison of the proposed method with S (sketching matrix size) and empirical baseline with N (MC trials). NRMSE (%) is measured against a high-accuracy reference map obtained with $N = 10{,}000$ (VORTEX on brain at $\alpha = 1$). For small sample sizes (≤ 50), Monte Carlo exhibits markedly higher error, while our method yields lower error at all examined sketch sizes. Even at $S = N = 100$, the proposed approach offers less error with an order-of-magnitude reduction in computational time (0.12s vs. 1.8s for MC).

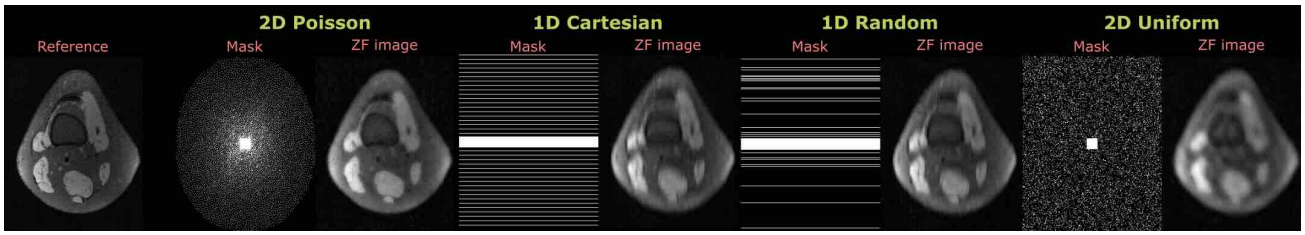

Figure 10. Different undersampling schemes analyzed in this study. For a representative slice, ZF images generated by the undersampling masks are showcased, along with the fully-sampled reference. Coherent and incoherent aliasing artifacts are generally exhibited by 1D and 2D undersampling masks, respectively.

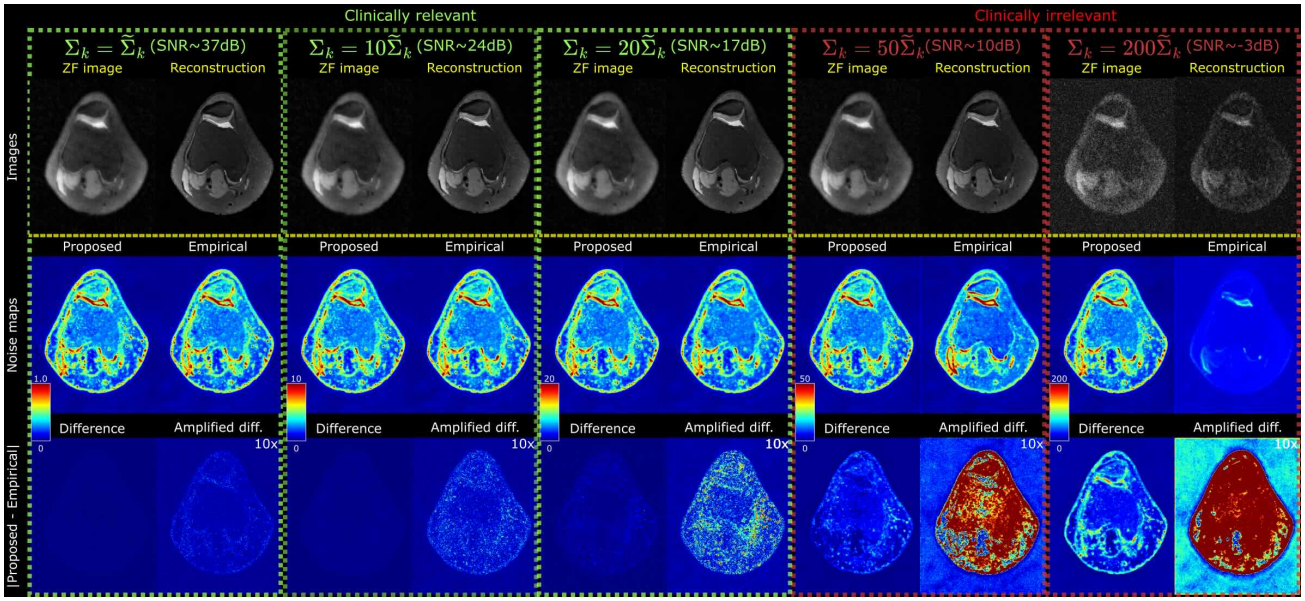

Figure 11. Each column corresponds to a different noise scaling factor α resulting in varying SNR scenarios (E2E-VarNet on knee, at $R = 8$). In each column: **(top row)** shows the zero-filled (ZF) image and the final reconstructed; **(middle row)** displays variance maps derived by the proposed method and empirical simulations; **(bottom row)** presents the difference map between the proposed and empirical variance estimates, plus an "amplified" difference to further highlight spatial discrepancies. The color bars indicate each map's display window in arbitrary units of variance. Note that $\alpha = 50$–200 corresponds to SNR values below $\approx 10\,\mathrm{dB}$ (i.e., well under an SNR of 10-15 that is often cited as a threshold for diagnostic utility.

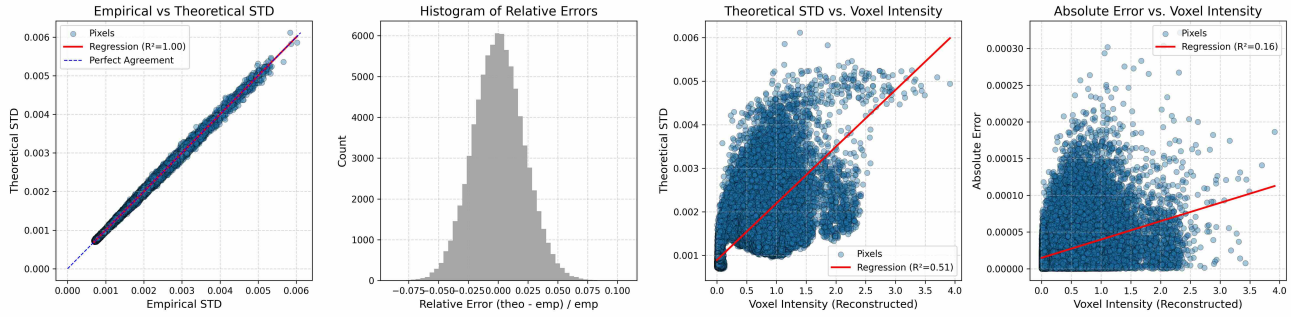

Figure 12. Additional visualizations for voxel-wise noise standard deviation (STD) analysis. (Left) Scatter plot of theoretical STD vs. empirical STD, showing high correlation ($R^2 \approx 1.00$). (Second from left) Histogram of relative errors between theoretical and empirical STD, indicating errors are centered around zero. (Third from left) Scatter plot of theoretical STD vs. reconstructed voxel intensity, showing a positive correlation ($R^2 \approx 0.51$). (Right) Scatter plot of absolute error in STD estimation vs. reconstructed voxel intensity, indicating a mild positive correlation ($R^2 \approx 0.16$).

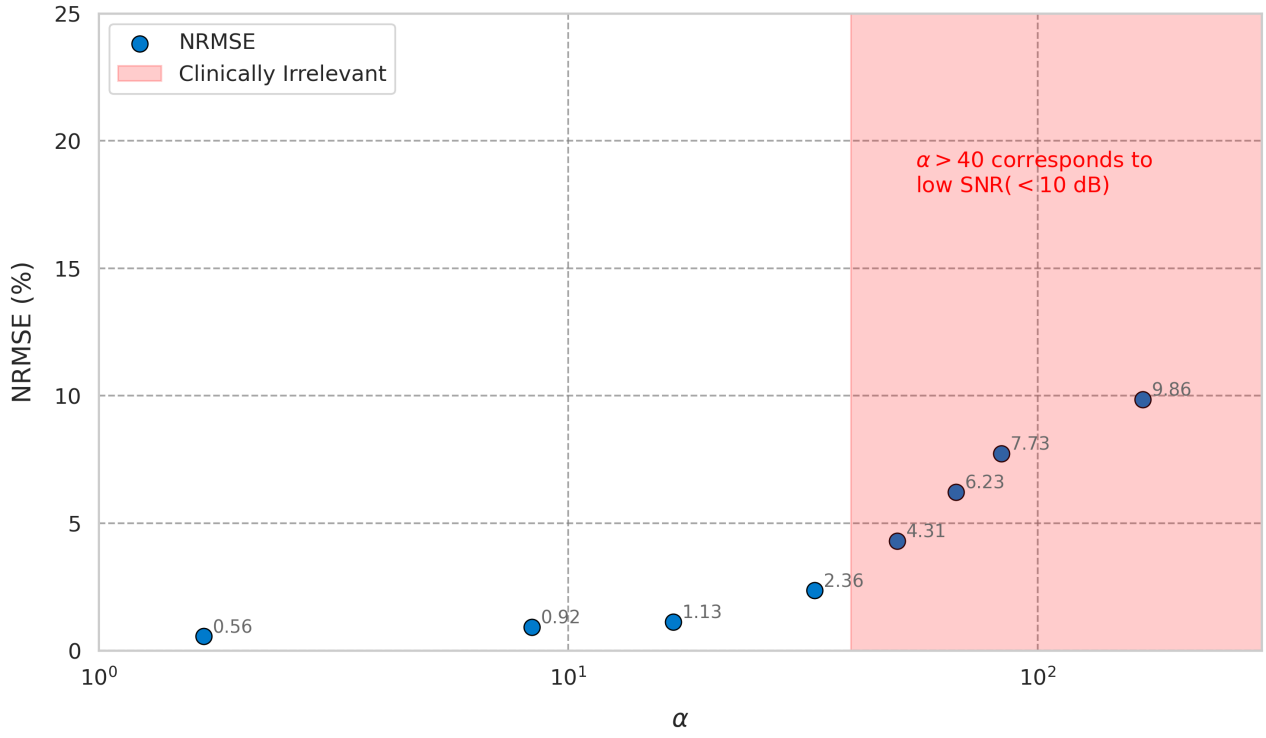

Figure 13. Mean NRMSE between reference empirical variance maps and variance maps from our method across test slices were plotted against increased noise levels (E2E-VarNet on knee, at $R = 8$). Here, α is plotted on a log scale and progressively scales the noise covariance matrix $\mathbf{\Sigma}_k = \alpha\widetilde{\mathbf{\Sigma}}_k$, simulating higher noise, and accordingly lower SNR levels.

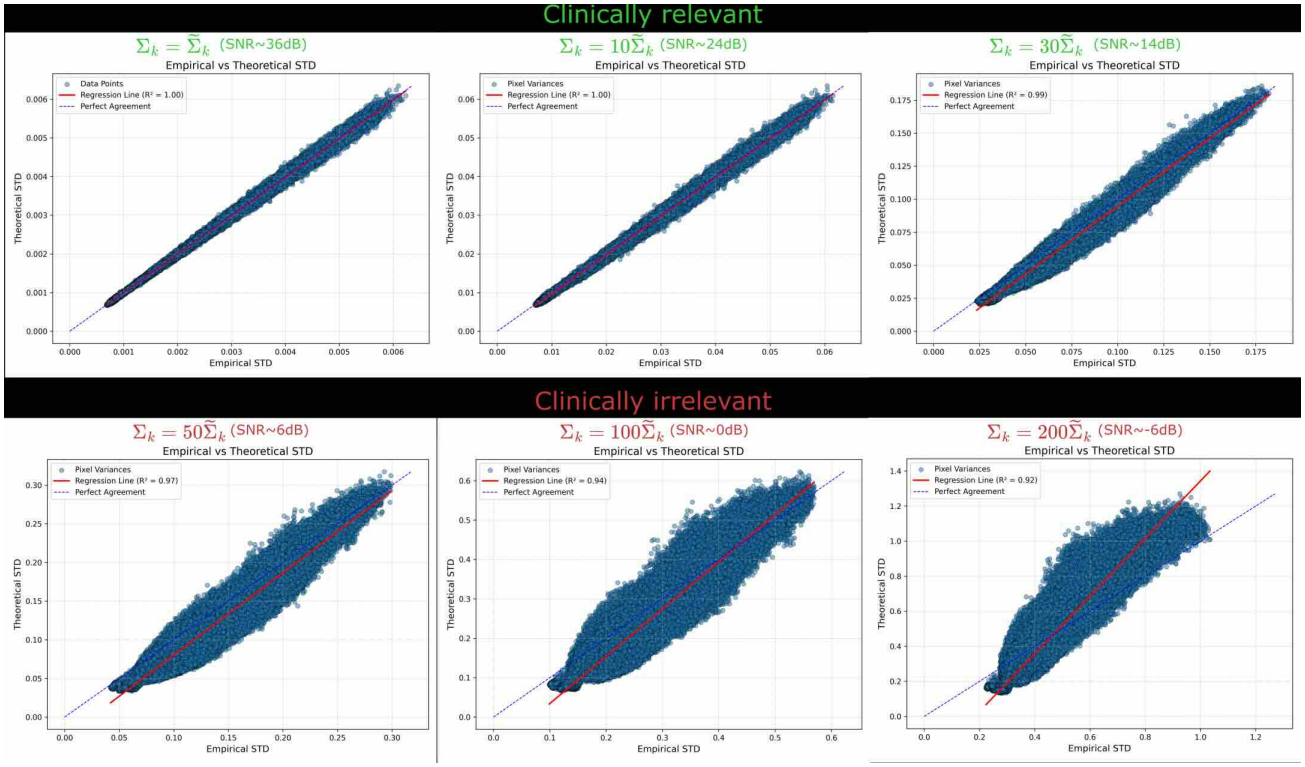

Figure 14. For a representative brain slice, empirical variances (horizontal axes) were compared to variances calculated our method (vertical axes) under successively larger noise scales: $\alpha = 1\,10, 30, 50, 100, 200$ (E2E-VarNet, at $R = 8$). In each panel, the diagonal line represents ideal one-to-one correspondence between empirical and calculated estimates, while the solid line indicates the best-fit linear regression (with its R^2 value indicated). Each data point corresponds to a single voxel's empirical vs. calculated variance, visualizing how closely they match at different noise levels.