Fast MRI for All: Bridging Access Gaps by Training without Raw Data

Yaşar Utku Alçalar
University of Minnesota
alcal029@umn.edu

Merve Gülle
University of Minnesota
glle0001@umn.edu

Mehmet Akçakaya*
University of Minnesota
akcakaya@umn.edu

Abstract

Physics-driven deep learning (PD-DL) approaches have become popular for improved reconstruction of fast magnetic resonance imaging (MRI) scans. Though PD-DL offers higher acceleration rates than existing clinical fast MRI techniques, their use has been limited outside specialized MRI centers. A key challenge is generalization to rare pathologies or different populations, noted in multiple studies, with fine-tuning on target populations suggested for improvement. However, current approaches for PD-DL training require access to raw k-space measurements, which is typically only available at specialized MRI centers that have research agreements for such data access. This is especially an issue for rural and under-resourced areas, where commercial MRI scanners only provide access to a final reconstructed image. To tackle these challenges, we propose **C**ompressibility-inspired **U**nsupervised Learning via **P**arallel **I**maging Fi**d**elity (CUPID) for high-quality PD-DL training using only routine clinical reconstructed images exported from an MRI scanner. CUPID evaluates output quality with a compressibility-based approach while ensuring that the output stays consistent with the clinical parallel imaging reconstruction through well-designed perturbations. Our results show CUPID achieves similar quality to established PD-DL training that requires k-space data while outperforming compressed sensing (CS) and diffusion-based generative methods. We further demonstrate its effectiveness in a zero-shot training setup for retrospectively and prospectively sub-sampled acquisitions, attesting to its minimal training burden. As an approach that radically deviates from existing strategies, CUPID presents an opportunity to provide broader access to fast MRI for remote and rural populations in an attempt to reduce the obstacles associated with this expensive imaging modality. Code is available at `https://github.com/ualcalar17/CUPID`.

1 Introduction

Magnetic resonance imaging (MRI) is a central tool in modern medicine, offering multiple soft tissue contrasts and high diagnostic sensitivity for numerous diseases. However, MRI is among the most expensive medical imaging modalities, in part due to its long scan times. Demand for MRI scans has shown an annual growth rate of 2.5%, while the number of MRI units per capita has increased by 1.8% in a similar time frame [55]. This mismatch has further increased the wait times for MRI exams [14, 42], particularly in rural areas and under-resourced/remote communities [17], shown in Fig. 1. Thus, techniques for fast MRI scans that can reduce overall scan times without compromising diagnostic quality [6] are critical for improving the throughput of MRI. Computational MRI approaches, including partial Fourier imaging [56], parallel imaging [61, 34], compressed sensing (CS) [53], and more recently deep learning (DL) [38, 63] have been developed for accelerating MRI.

*Corresponding Author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

In most MRI centers, parallel imaging remains the most widely used approach for the reconstruction of routine clinical images. The acceleration rates afforded by these methods, however, are limited due to noise amplification and aliasing artifacts. DL-based methods, especially physics-driven DL (PD-DL) approaches, offer state-of-the-art improvements over parallel imaging [48]. However, the translation of PD-DL to clinic has been hindered by generalizability and artifact issues related to details not well represented in training databases, in other words when faced with out-of-distribution samples at test time [31, 49, 59, 11]. This is a problem for many typical MRI centers, whose population characteristics do not necessarily align with specialized MRI centers in urban settings, where training databases for PD-DL are currently curated.

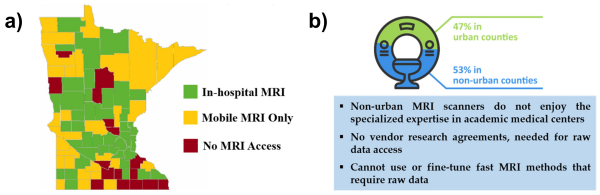


Figure 1: Many regions lack direct MRI access or rely on local/mobile units: (a) Over half of MRI services in Minnesota are in non-urban areas [17], and (b) these scanners often lack vendor agreements for raw data access, limiting AI fine-tuning.

In such cases, fine-tuning of the PD-DL model on the target population may be beneficial [47, 25, 74, 19]. However, a major roadblock for this strategy is that all current training methods for PD-DL require access to raw MRI data. Such access requires research agreements with MR vendors, and is typically not available outside specialized/academic MRI centers. This is especially an issue for rural and under-resourced areas, where commercial MRI scanners only provide access to a final image, reconstructed via parallel imaging.

In this work, we tackle these challenges associated with typical PD-DL training, and propose **C**ompressibility-inspired **U**nsupervised Learning via **P**arallel **I**maging Fi**d**elity (CUPID), which trains PD-DL reconstruction from routine clinical images, *e.g.*, in Digital Imaging and Communications in Medicine (DICOM) format. Succinctly, CUPID uses a compressibility-inspired term to evaluate the goodness of the output, while ensuring the output is consistent with parallel imaging via well-designed input perturbations. CUPID can be used both with database-training and in a subject-specific/zero-shot manner, attesting to its minimal fine-tuning burden. Our key contributions include:

- We introduce CUPID, a novel method that enables high-quality training of PD-DL reconstruction in unsupervised and zero-shot/subject-specific settings. By using only routine clinical reconstructed MR images, CUPID eliminates the need for access to raw k-space measurements. To the best of our knowledge, our method is the first attempt to train PD-DL networks using only these images that are exported from the scanner.
- CUPID is trained on DICOM images acquired at the *target acceleration rate*, which often have noise and aliasing artifacts due to high sub-sampling, in an unsupervised manner. Note this is a deviation from other methods that use reference fully-sampled DICOM images to train a likelihood or score function, such as generative models.
- CUPID uses a novel unsupervised loss formulation that enforces fidelity with using parallel imaging algorithms via carefully designed perturbations, in addition to evaluating the compressibility of the output image. This parallel imaging fidelity ensures the network does not converge to overly sparse solutions.
- We provide a comprehensive evaluation encompassing both retrospective and prospective acquisitions at target acceleration rates, along with expert radiologist assessments and pathology cases from FastMRI+ [79], demonstrating that CUPID achieves reconstruction quality comparable to leading supervised and self-supervised methods that require raw k-space data, while outperforming conventional CS techniques and diffusion-based generative methods.

2 Background and Related Work

2.1 MRI forward model and conventional methods for MRI reconstruction

MRI raw data is acquired in the frequency domain of the image, referred to as k-space. For fast MRI, data is acquired in the sub-Nyquist regime by undersampling the acquisition in k-space. In this

case, the forward acquisition model relating the image $\mathbf{x} \in \mathbb{C}^n$ to these raw MRI data (or k-space) measurements is given as:

$$\mathbf{y}_\Omega = \mathbf{E}_\Omega \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}_\Omega$ denotes the acquired k-space data corresponding to the undersampling pattern Ω with $|\Omega| = m < n$. $\mathbf{E}_\Omega$ denotes the multi-coil encoding operator that includes information from n_c receiver coils, each of which are sensitive to a different part of the image [37]. When the acceleration rate $R = n/m$ is less than n_c, this system of equations is over-determined due to the redundancies among the receiver coils. Parallel imaging uses these redundancies to solve the maximum likelihood estimation problem under i.i.d. Gaussian noise [61]:

$$\mathbf{x}_{\text{PI}} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 = (\mathbf{E}_\Omega^H \mathbf{E}_\Omega)^{-1} \mathbf{E}_\Omega^H \mathbf{y}_\Omega. \tag{2}$$

Numerically, this can be solved directly for certain undersampling patterns [61] or more broadly iteratively using conjugate gradient (CG) [60]. Using the equivalence of multiplication in image domain and convolutions in k-space [68], it can also be solved as an interpolation problem in k-space [34]. Parallel imaging remains the most clinically used acceleration method for MRI, with some MR systems using the image-based reconstruction, while others utilizing the equivalent k-space interpolation.

In modern computational MRI, additional regularization is often incorporated into the objective function [39]:

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{3}$$

where $\mathcal{R}(\cdot)$ denotes a regularizer. For instance, CS uses the idea that images should be compressible in an appropriate transform domain [53], and uses $\mathcal{R}(\mathbf{x}) = \tau ||\mathbf{W}\mathbf{x}||_1$, where τ is the regularization weight, $\mathbf{W}$ is a linear sparsifying transform such as a discrete wavelet transform (DWT) and $||\cdot||_1$ is the ℓ_1 norm.

2.2 PD-DL reconstruction via algorithm unrolling

Among different PD-DL methods [3, 33, 48], unrolled networks [58] remain the highest performer in reconstruction challenges [39, 59]. These methods unroll iterative algorithms for solving the regularized least squares objective in Eq. (3) [32], such as proximal gradient descent [63, 43] or variable splitting with quadratic penalty (VS-QP) [2], over a fixed number of steps. Unrolled networks are conventionally trained using supervised learning over a database, where the reference raw k-space measurements are first retrospectively undersampled to form $\mathbf{y}_\Omega$. Subsequently, the network is trained to map to the original full reference k-space or the corresponding reference image [38, 2, 77] by minimizing:

$$\min_{\boldsymbol{\theta}} \mathbb{E}\left[\mathcal{L}\left(\mathbf{y}_{\text{ref}}, \mathbf{E}_{\text{full}}(f(\mathbf{y}_\Omega, \mathbf{E}_\Omega; \boldsymbol{\theta}))\right)\right] \tag{4}$$

where $\boldsymbol{\theta}$ are the network parameters, $f(\mathbf{y}_\Omega, \mathbf{E}_\Omega; \boldsymbol{\theta})$ denotes the network output for inputs $\mathbf{y}_\Omega$ and $\mathbf{E}_\Omega$, $\mathbf{E}_{\text{full}}$ is the fully-sampled encoding operator, $\mathbf{y}_{\text{ref}}$ is the fully-sampled reference k-space data, and $\mathcal{L}(\cdot, \cdot)$ is a loss function.

2.3 Self-supervised and unsupervised methods

Obtaining fully-sampled reference data in MRI can be infeasible due to prolonged scan durations, organ movement in acquisitions such as real-time cardiac imaging or myocardial perfusion [62], or signal decay in acquisitions like diffusion MRI with EPI [69]. To enable training of PD-DL networks without fully sampled raw MRI data, a variety of unsupervised learning methodologies have emerged [5], including self-supervised learning techniques [75, 20] and generative modeling approaches [45, 23, 22].

Self-supervised methods generate supervisory labels from the undersampled data itself [75, 20, 57, 44]. A pioneering method in this field, self-supervision via data undersampling (SSDU) [75, 73], involves partitioning the acquired measurement indices Ω into two disjoint subsets ($\Omega = \Lambda \cup \Theta$) to train the network in a self-supervised manner:

$$\min_{\boldsymbol{\theta}} \mathbb{E}\left[\mathcal{L}\left(\mathbf{y}_\Lambda, \mathbf{E}_\Lambda(f(\mathbf{y}_\Theta, \mathbf{E}_\Theta; \boldsymbol{\theta}))\right)\right] \tag{5}$$

Even though these self-supervision based approaches demonstrate exceptional performance across various tasks, they lack the ability to train the model without access to undersampled raw data, as they cannot operate solely using images that are exported from the scanner.

Conversely, generative methods learn the prior distribution of the given dataset, which is then leveraged in conjunction with a log-likelihood data term during the testing phase. Although recent methods based on diffusion/score-based models have shown substantial promise, these methods require large amounts of high-quality images either reconstructed from raw data [45, 52] or as DICOMs [23], as well as computational resources to perform the training, both of which may not be feasible in the setups we are focused on.

3 Methodology

3.1 Why is it important to train PD-DL networks without raw k-space data?

A major limitation in the clinical adoption of deep learning (DL)-based MRI reconstruction is the issue of generalizability [38, 46, 47, 59, 26, 41, 80]. Models trained on data from a specific scanner, patient population, or acquisition protocol often fail when applied to different settings due to distribution shifts [59, 51]. This lack of robustness can lead to artifacts or hallucinations which are false structures that resemble real anatomy, but are not present in the underlying data [47, 59, 41, 51].

Studies have shown that pre-trained models struggle to adapt to real-world variations, particularly in centers where patient demographics, scanner configurations, or imaging protocols differ from the original training data [41]. This limitation was highlighted in the FastMRI challenges in the transfer track [59], where networks trained on images from one vendor failed to generalize to data from another. Traditional solutions involve fine-tuning models with additional raw k-space data from the target domain [27, 74], but this is infeasible in many clinical environments. While all MRI scanners inherently sample k-space data, the majority of scanners outside specialized academic or research institutions (*e.g.*, as those in local hospitals and mobile MRI units) lack the ability to export this data due to vendor restrictions [72]. As a result, most DL-based reconstruction methods, which typically rely on raw k-space during training or fine-tuning, cannot be applied in these settings.

Outside the context of PD-DL methods, diffusion models as generative priors offer an alternative in this setting. Initial promising results [45, 23, 66] have suggested more robustness to sampling pattern shifts, and some robustness to anatomy changes, though more recent works have argued that diffusion priors also face generalizability issues for out-of-distribution samples [13]. Thus, while promising, the robustness of the learned prior to distribution shifts, as well as the heuristic tuning of data fidelity parameters during posterior sampling, require further investigation. Furthermore, diffusion models are trained on either raw data [45, 1], fully-sampled magnitude [23] or complex-valued [22, 10, 36] DICOMs. However, they have not been trained on sub-sampled DICOMs that still contain artifacts, as explored in this work. Finally, there are other challenges for the clinical applicability and acceptance of diffusion models in MRI, including the longer inference times compared to PD-DL, and more importantly the stochasticity of the output.

In light of these, training/fine-tuning PD-DL networks without raw k-space data is not just a matter of convenience, but a necessary step toward ensuring that DL-based MRI reconstruction can generalize across diverse clinical settings.

3.2 Training without raw k-space data

In this study, we introduce a novel framework to train PD-DL models, utilizing only routinely available clinical images exported directly from MRI scanners. Recently, inspired by the connections between PD-DL and compressibility-based processing [35], a compressibility-inspired loss was proposed to evaluate the goodness of unsupervised PD-DL training [9]. However, this approach still requires access to raw k-space data to stabilize training, making it unsuitable for our goals. Here, we adapt the compressibility idea and augment it with a parallel imaging fidelity to successfully reconstruct clinical images in DICOM format without needing any raw k-space data.

Compressibility aspect of the loss formulation Compressibility/sparsity in the output of the PD-DL network can be enforced by utilizing a weighted ℓ_1 norm [9], which has been demonstrated

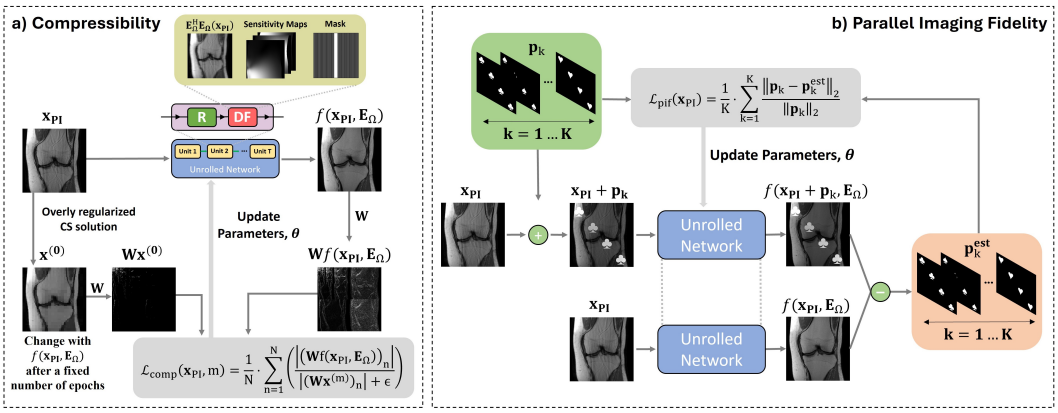


Figure 2: Our Compressibility-inspired Unsupervised Learning via Parallel Imaging Fidelity (CUPID) method trains PD-DL models in an unsupervised and/or zero-shot manner without requiring any raw k-space data. The network is unrolled for T units, with each unit consisting of regularizer (R) and data fidelity (DF). The proposed loss function comprises two terms: (a) a reweighted ℓ_1 component that assesses the compressibility of the network's output; (b) a fidelity term that ensures the output stays consistent with parallel imaging reconstructions via carefully designed perturbations, thereby preventing the network from producing a sparse all-zeros output.

to provide a closer approximation to the ℓ_0 norm compared to the standard ℓ_1 norm [18]. Thus, this compressibility of the output image in CUPID is achieved by the loss term

$$\mathcal{L}_{\text{comp}}(\mathbf{x}_{\text{PI}}, \text{m}) = \frac{1}{N} \cdot \sum_{n=1}^{N} \left(\frac{|(\mathbf{W} f(\mathbf{x}_{\text{PI}}, \mathbf{E}_\Omega))_n|}{|(\mathbf{W}\mathbf{x}^{(\text{m})})_n| + \epsilon} \right), \tag{6}$$

where $\mathbf{x}_{\text{PI}}$ denotes the DICOM input acquired using parallel imaging, $\mathbf{W}$ represents the wavelet transform, N is the total number of wavelet coefficients and $\mathbf{x}^{(\text{m})}$ signifies the signal estimate following the training during the m^{th} reweighting step. Similar to [9, 18], we chose the initial weights from a CS reconstruction that has a large regularization and ϵ is added for numerical stability. Note, here we redefined $f(\cdot, \cdot)$ without the network parameters, $\boldsymbol{\theta}$, and used $\mathbf{x}_{\text{PI}}$ as the network input instead of $\mathbf{y}_{\boldsymbol{\Omega}}$, to simplify notation.

Parallel imaging fidelity via perturbation equivariance Relying solely on Eq. (6) will result in inaccurate training as the neural network learns to produce an all-zeros image in an effort to drive the wavelet coefficients in the numerator to zero, which minimizes the loss function in Eq. (6). In [9], fidelity with raw k-space data was used to avoid this training issue. In our setting without raw k-space access, we introduce a novel fidelity operator that stabilizes the training of the reconstruction algorithm, building on ideas from parallel imaging. Specifically, we define a secondary loss term that enforces parallel imaging fidelity formulated as:

$$\mathcal{L}_{\text{pif}}(\mathbf{x}_{\text{PI}}) = \mathbb{E}_{\mathbf{p}}\big[||\mathbf{p} - \mathbf{p}^{\text{est}}||_2 / ||\mathbf{p}||_2\big], \tag{7}$$

where $\mathbf{p}^{\text{est}} = f(\mathbf{x}_{\text{PI}} + \mathbf{p}, \mathbf{E}_\Omega) - f(\mathbf{x}_{\text{PI}}, \mathbf{E}_\Omega)$. To motivate this loss term, we build upon the theoretical foundation of equivariant imaging (EI) [20]. In EI, an equivariance property for the composition of the forward operator and the reconstruction network is assumed with respect to a transformation group, and necessary conditions are derived for signal recovery [20].

Our work adapts this framework by introducing a novel transformation group tailored to parallel imaging (PI) in MRI. We note that parallel imaging reconstructions exhibit an equivariance property with respect to a specific group of perturbations $\mathcal{P} = \{\mathbf{p}_k\}_{k=1}^{K}$, which are designed to ensure that their R-fold aliasing do not create overlaps in the field-of-view. For $\mathbf{p} \in \mathcal{P}$, we let $T_{\mathbf{p}}(\mathbf{x}) = \mathbf{x} + \mathbf{p}$ be the affine perturbation from this group. Our parallel imaging fidelity then assumes equivariance of the reconstruction network to perturbations from this group as follows:

$$f(T_{\mathbf{p}}(\mathbf{x}_{\text{PI}}), \mathbf{E}_\Omega) = f(\mathbf{x}_{\text{PI}} + \mathbf{p}, \mathbf{E}_\Omega) = f(\mathbf{x}_{\text{PI}}, \mathbf{E}_\Omega) + \mathbf{p} = T_{\mathbf{p}}\big(f(\mathbf{x}_{\text{PI}}, \mathbf{E}_\Omega)\big), \tag{8}$$

Our second loss term, defined in Eq. (7), promotes this equivariant behavior by penalizing deviations from this property. We also note this set of perturbations satisfy the necessary conditions in [20]. In

doing so, it implicitly encodes parallel imaging consistency without relying on access to raw k-space data. Our final loss function for CUPID is:

$$\mathcal{L}_{\text{CUPID}} = \mathcal{L}_{\text{comp}} + \lambda \cdot \mathcal{L}_{\text{pif}}, \tag{9}$$

where λ is a trade-off parameter between two terms.

Perturbation design specifics The perturbations used for R-fold acceleration are constructed to avoid aliasing overlaps in the field-of-view, ensuring they are resolvable via parallel imaging. This design aligns with the equivariance assumption introduced above: the network should recover an accurate estimate of $\mathbf{x}$ from $\mathbf{x}_{\text{PI}}$, and $\mathbf{x} + \mathbf{p}$ from $\mathbf{x}_{\text{PI}} + \mathbf{p}$. Both cases are illustrated in Fig. 2.

From an implementation perspective, the expectation over $\mathbf{p}$ is calculated over K such perturbations $\{\mathbf{p}_k\}$. The fold-over constraint for each $\{\mathbf{p}_k\}$ is achieved by picking the perturbations as randomly rotated and positioned letters, numbers, card suits or other shapes that have different intensity values. These choices also ensure that high-frequency information, such as edges, are accurately reconstructed by the regularization process. Further information on this is given in Appendix C.2.

Subject-specific/zero-shot application In resource-limited settings, it may be more practical to fine-tune the method using only a few subjects, or even a single subject, to significantly reduce computational costs. As Eq. (6) does not solely focus on the subtraction between two entities, it lacks an inherent mechanism to drive the loss to zero through overfitting. Therefore, CUPID can be tailored to suit a scan-specific context [4] without any modification to the loss given in Eq. (9).

4 Evaluation

4.1 Experimental setup and implementation details

We conducted a thorough evaluation of our method, assessing its performance through both qualitative and quantitative analyses, and focused on uniform/equidistant patterns which produces coherent artifacts that are more difficult to remove compared to the incoherent artifacts from random undersampling [47].

Retrospective undersampling setup In our retrospective studies, we used fully-sampled multi-coil knee and brain MRI data from the fastMRI database, acquired with relevant institutional review board approvals [49, 78]. Knee dataset included fully-sampled coronal proton density-weighted (coronal PD) and PD with fat suppression (coronal PD-FS) data, both with matrix sizes of 320×320. For brain MRI, axial T2-weighted (ax T2) and axial FLAIR (ax FLAIR) datasets with matrix size of 320×320 are used. Both knee datasets comprise data collected from 15 receiver coils, while the brain datasets include data collected from 16 and 20 receiver coils for the ax T2 and FLAIR, respectively.

Every dataset was retrospectively undersampled using a uniform/equidistant pattern at $R = 4$. 24 lines of auto-calibration signal (ACS) from center of the raw k-space data were kept. DICOM images to train our proposed model were reconstructed using parallel imaging (CG-SENSE), solving $\mathbf{x}_{\text{PI}} = (\mathbf{E}_\Omega^H \mathbf{E}_\Omega)^{-1} \mathbf{E}_\Omega^H \mathbf{y}_\Omega$. For each dataset, models were trained using 300 slices, and testing was performed using 380 slices for knee MRI and 300 slices for brain MRI, from distinct subjects. More details about the implementation of the PD-DL models are provided in Appendix A.

Prospective undersampling setup In this experiment, we replicate the practical pipeline for CUPID, where data is acquired at the desired high acceleration rate, and reconstructed to $\mathbf{x}_{\text{PI}}$ with noise and aliasing artifacts, using parallel imaging. To this end, a 3D MP2RAGE sequence [54] was acquired on a 7T Siemens Magnetom MRI scanner, with institutional review board approval, and matrix size$= 320 \times 300 \times 224$, isotropic resolution $= 0.75$mm, prospective undersampling $R = 4$ (in k_y only). This is the desired target acceleration, which is higher than the standard clinical acquisition. Low-resolution images were acquired in the same orientation for sensitivity estimation [50]. Training and reconstruction with CUPID was done in a zero-shot subject-specific manner.

4.2 Comparison methods

We compared our method with several database training methods that have access to raw k-space data, including supervised PD-DL [38, 2, 48], SSDU [75], and equivariant imaging (EI) [20]. All

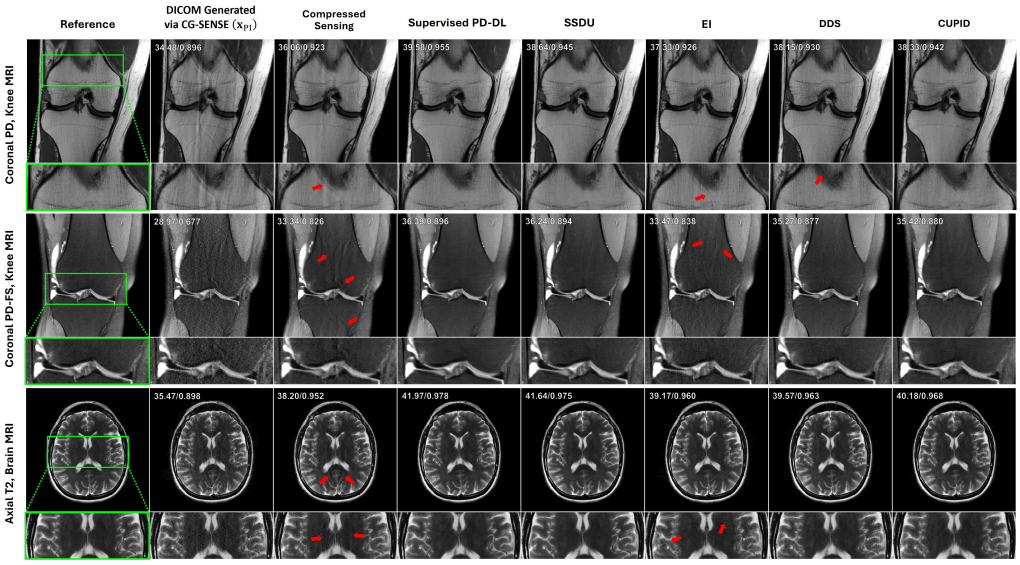


Figure 3: Representative slices reconstructed at an acceleration factor of $R = 4$ using equidistant undersampling from coronal PD and coronal PD-FS knee MRI, as well as axial T2-weighted brain MRI. The baseline CG-SENSE, CS, EI-trained PD-DL, and DDS suffer from residual artifacts highlighted by red arrows. PD-DL trained with CUPID improves upon them while delivering reconstruction quality comparable to supervised and SSDU-trained PD-DL.

PD-DL methods used the same unrolled network and components (Appendix A) to ensure that only the training process differed for fair comparisons. We also compared CUPID with recent diffusion model based techniques, ScoreMRI [23] and DDS [22]. These train a time-dependent score function using denoising score matching on a large dataset of reference fully-sampled DICOM images or raw data, and uses this score function during inference to sample from the conditional distribution given the measurements. Finally, we also included conventional reconstruction methods, CG-SENSE, which was used to generate the original $\mathbf{x}_{\text{PI}}$ as the clinical baseline comparison that is typically not clinically usable at high acceleration rates [38], as well as CS [53].

We note that all aforementioned methods use $\mathbf{E}_\Omega^H \mathbf{y}_\Omega$ for data fidelity during inference, since this is what is needed to run gradient descent or conjugate gradient on ℓ_2-based fidelity terms including $\mathbf{y}_\Omega$ and $\mathbf{E}_\Omega \mathbf{x}$, as seen in Eq. (2) or SuppMat Sec. 6. $\mathbf{E}_\Omega^H \mathbf{y}_\Omega$ can be formed without accessing $\mathbf{y}_\Omega$ by simply multiplying $\mathbf{x}_{\text{PI}}$ with $\mathbf{E}_\Omega^H \mathbf{E}_\Omega$. Note that $\mathbf{E}_\Omega$ includes information about the undersampling pattern Ω, which is completely known from the acquisition parameters, and coil sensitivities, which can be estimated from separate calibration scans in DICOM format [50]. We emphasize that what sets CUPID apart from other PD-DL strategies is that it is the only one that can train the unrolled network without using $\mathbf{y}_\Omega$. Thus, without loss of generality, $\mathbf{E}_\Omega$ is known both in training and in testing for all methods.

In the zero-shot setup, we compared our zero-shot results with zero-shot SSDU (ZS-SSDU) [74] as well as CS, and our baseline method, CG-SENSE - all of which are compatible with zero-shot inference. We also include DDS and ScoreMRI for inference in this setup, while noting that these diffusion priors have been pre-trained on very large databases. This is done to highlight CUPID's performance and versatility, which is trained in a zero-shot manner on the given slice. All quantitative evaluations used structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). We further emphasize that CUPID does not use raw k-space data, and as a result, it does not benefit from the redundancies across multiple coils due to the lack of access to multi-coil raw k-space data. Therefore, resources are further restricted, unlike the comparison methods.

4.3 Experiments with retrospective undersampling

Database results Representative results in Fig. 3 show that baseline CG-SENSE, CS and EI reconstructions exhibit residual artifacts, whereas DDS suffers from minor noise amplification. In

Table 1: Quantitative results for comparison methods on Coronal PD, Coronal PD-FS, Ax FLAIR, and Ax T2 datasets using equispaced undersampling pattern at $R = 4$. First 3 rows: Gold standard setup. Trained on the same organ, contrast and acceleration rate as the target population using fully-sampled or sub-sampled raw data. Mid 2 rows: Trained on the same organ as the target population, using either fully-sampled raw data or fully-sampled artifact-free DICOM images. Last 3 rows: No access to raw data; only sub-sampled DICOM images with artifacts. The **best** and second-best values are highlighted, excluding the gold standard setup.

Method	Cor PD, Knee MRI		Cor PD-FS, Knee MRI		Ax FLAIR, Brain MRI		Ax T2, Brain MRI	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Supervised [38]	40.44	0.964	35.72	0.893	37.91	0.967	36.77	0.933
SSDU [75]	39.64	0.957	35.68	0.892	37.55	0.964	36.59	0.931
EI [20]	38.07	0.938	33.83	0.849	36.40	0.953	34.58	0.909
ScoreMRI [23]	37.85	0.928	35.01	**0.883**	35.26	0.934	33.83	0.899
DDS [22]	38.64	0.950	34.89	0.875	36.18	0.950	35.16	0.914
PI [60, 34]	35.51	0.909	29.48	0.735	31.97	0.906	31.02	0.833
CS [53]	36.92	0.922	33.31	0.841	33.17	0.929	33.61	0.897
CUPID **(ours)**	**38.82**	**0.952**	**35.04**	0.880	**36.49**	**0.957**	**35.31**	**0.921**

contrast, CUPID successfully eliminates these artifacts from the CG-SENSE image using a well-trained PD-DL network, achieving a state-of-the-art reconstruction quality comparable to supervised PD-DL and SSDU, despite only having access to $\mathbf{x}_{\text{PI}}$ for training, and not to raw k-space data unlike these other methods. We observe that parallel imaging reconstruction is not clinically usable at higher acceleration rates, but is improved using a CUPID-trained PD-DL reconstruction. Quantitative results in Tab. 1 support visual observations, showing that CUPID consistently outperforms CG-SENSE, CS, and EI across multiple datasets. It also outperforms ScoreMRI and DDS across various scenarios in the majority of cases, while achieving performance comparable to supervised PD-DL and SSDU, both of which have access to raw data. Further qualitative results are provided in Appendix I.

Subject-specific/zero-shot learning results Fig. 4 shows zero-shot reconstruction results. CG-SENSE and CS suffer from noise amplification and residual artifacts, with CG-SENSE showing more severe degradation. ScoreMRI and DDS exhibit artifacts in coronal PD, and ScoreMRI introduces blurring in coronal PD-FS. This blurring leads to higher distortion metrics (*e.g.*, PSNR/SSIM), which favor smoother outputs due to penalties on high-frequency details, reflecting the perception-distortion

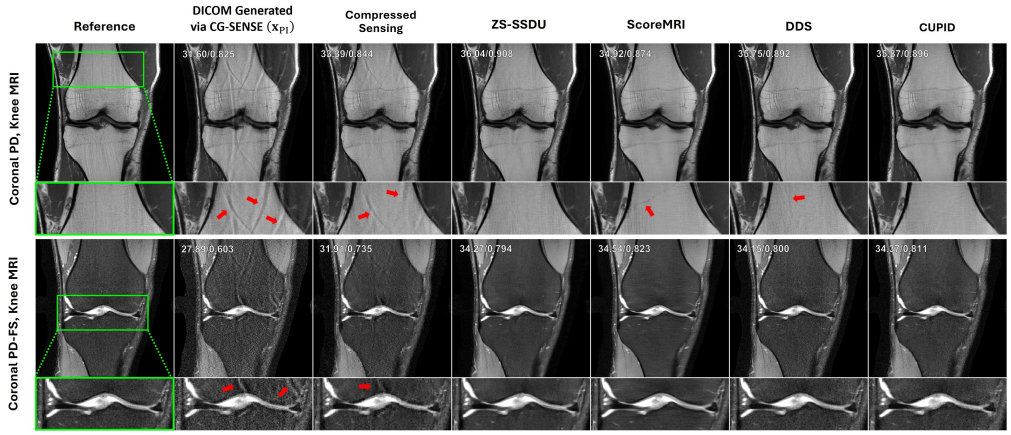


Figure 4: Representative subject-specific/zero-shot learning results for various algorithms on coronal PD and coronal PD-FS knee MRI for retrospective $R = 4$ equidistant undersampling, along with results from database-trained diffusion models on knee data. Baseline CG-SENSE, CS, ScoreMRI and DDS suffer from residual artifacts (red arrows) and blurring. PD-DL with CUPID loss successfully removes these artifacts, and functions in a similar manner to ZS-SSDU.

trade-off [16, 21, 7]. CUPID again achieves superior artifact and noise reduction, closely matching ZS-SSDU quality despite lacking raw data or self-validation.

Ablation studies We further conducted three ablation studies to examine key factors influencing CUPID's performance. The first examined the effect of the number of perturbations (Appendix C.1), the second explored the impact of the hyperparameter λ across multiple values (Appendix E), and the third evaluated CUPID's robustness to different sampling patterns and higher acceleration factors (Appendix F).

4.4 Practical setting: Prospective undersampling

As discussed in Section 4.1, brain data is acquired at the target acceleration rate, reconstructed via parallel imaging and exported in DICOM format for zero-shot fine-tuning. Fig. 5 shows reconstruction results for the vendor parallel imaging reconstruction, as well as CS, DDS and CUPID. Vendor-provided $R = 4$ reconstruction on the scanner exhibits residual artifacts in Fig. 5a (shown in zoomed insets). CS effectively reduces noise but results in an overly-smooth reconstruction. DDS slightly mitigates the artifact and reduces noise, but both remain present. Our proposed CUPID method delivers the most effective artifact and noise mitigation in the DICOM image without requiring any raw k-space data, demonstrating its real-world effectiveness. We note that ZS-SSDU cannot be applied here due to the unavailability of raw data. We also note that the vendor-provided DICOM was generated using k-space interpolation [34] instead of the image domain formulation in Eq. (2). Due to their equivalence, this did not cause any issues for CUPID, as expected. Furthermore, vendor-provided DICOMs may also include additional zero-padding [65], which our physics-based approach handles naturally, as well as additional filtering/processing. More information related on these aspects is given in Appendix D and Appendix G.

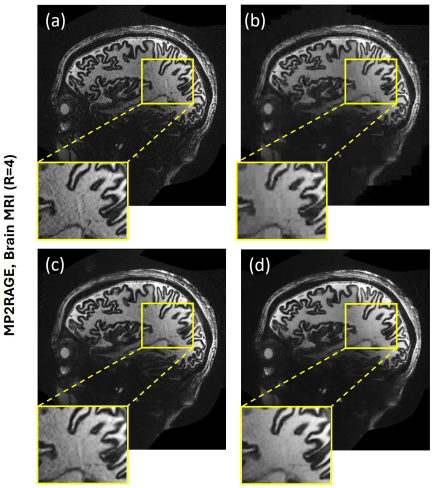


Figure 5: Prospective acceleration results for various methods that operate on parallel imaging-reconstructed DICOMs exported from the scanner. (a) Vendor-provided PI DICOM ($R = 4$), (b) CS, (c) DDS, and (d) CUPID.

4.5 Out-of-distribution (OOD) example

We further illustrate the generalizability issues discussed in Section 3.1 with a representative out-of-distribution example, intended to emulate the transfer from urban training datasets to inference in a rural setting, where scanner hardware and acquisition protocols often differ. Specifically, in Fig. 6, we demonstrate how SSDU and supervised PD-DL models trained on 3T coronal PD data struggle when applied to 1.5T data with a different SNR, leading to noticeable artifacts. DDS, trained on both 3T and 1.5T data, does a better job of mitigating this issue, though some performance degradation persists. CUPID, zero-shot trained on the given slice without the need for raw data, performs the best, effectively reducing artifacts and maintaining high image quality.

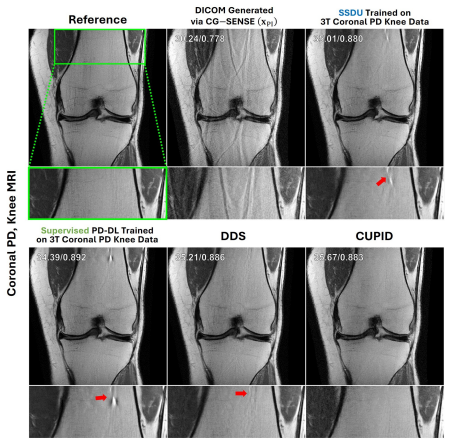


Figure 6: 1.5T OOD example showing CUPID's superior artifact reduction.

4.6 Qualitative radiologist readings including FastMRI+ pathology cases

To assess the clinical plausibility of CUPID's results, we consulted a musculoskeletal radiologist and a neuroradiologist, whose areas of expertise align with the knee and brain datasets. Both experts

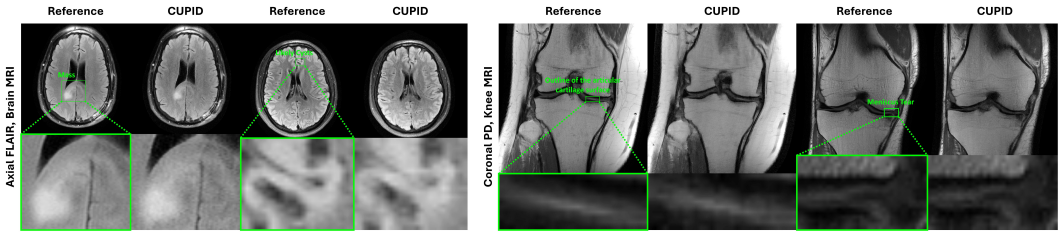


Figure 7: Representative pathology cases from the FastMRI+ dataset [79] showing CUPID reconstructions alongside reference images for brain and knee pathologies. Qualitative expert radiologist assessments suggest that CUPID maintains diagnostic content and visual fidelity across pathology regions of interest.

independently and blindly reviewed the reconstructed images presented in this work, as well as representative annotated pathology cases from the FastMRI+ dataset [79], and provided an evaluation of diagnostically relevant contents and overall clinical realism.

In retrospective experiments regarding database results (Fig. 3), the musculoskeletal radiologist observed that CUPID led to slightly more blurring than the supervised and self-supervised PD-DL methods, though this was not deemed diagnostically significant. Furthermore, CUPID was judged superior to DDS for noise suppression and artifact reduction by both radiologists. For the zero-shot experiments (Fig. 4), the musculoskeletal radiologist reported no noticeable blurring relative to ZS-SSDU and again found CUPID to yield lower noise than DDS.

For the prospective reconstructions (Fig. 5), reconstructions (C) and (D) were favored for their lower noise levels, whereas (A) was identified as the noisiest. The neuroradiologist preferred (D) overall, citing clearer gray–white matter boundaries and less pixelation compared to the other variants.

Finally, in the pathology evaluation (Fig. 7), representative FastMRI+ cases were reviewed by the same sub-specialists. For the brain dataset, the neuroradiologist compared CUPID and reference reconstructions of a mass and a likely cyst, noting no visible differences in the pathology regions and confirming that diagnostic information was maintained. For the knee dataset, the musculoskeletal radiologist assessed meniscus tear (complex degenerative tearing of the body of the meniscus with meniscal extrusion) and detailed depiction of the articular surface, which is critical for evaluating cartilage integrity. In both cases, CUPID reconstructions were considered diagnostically similar to the references in the clinically relevant areas.

5 Conclusion

In this study, we proposed Compressibility-inspired Unsupervised Learning via Parallel Imaging Fidelity (CUPID), a novel training strategy for PD-DL MRI reconstruction using only clinically accessible images, without raw k-space data. CUPID leverages image compressibility and carefully designed perturbations that preserve parallel imaging consistency, enabling high-quality, physics-driven reconstructions. To our knowledge, this is the first method to train PD-DL networks solely from clinical images. CUPID also alleviates the training burden of generative methods, which requires a large number of data during training to capture the prior well. Extensive retrospective and prospective evaluations demonstrate the effectiveness of CUPID across diverse MRI scans and learning settings.

Acknowledgements The authors gratefully acknowledge Dr. Jutta Ellermann and Dr. Can Özütemiz for providing expert radiologist assessments during the rebuttal stage on short notice.

References

[1] A. Aali, G. Daras, B. Levac, S. Kumar, A. Dimakis, and J. Tamir. Ambient diffusion posterior sampling: Solving inverse problems with diffusion models trained on corrupted data. In *Proc. Int. Conf. Learn. Represent.*, 2025.

[2] H. K. Aggarwal, M. P. Mani, and M. Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans. Med. Imag.*, 38(2):394–405, 2019.

[3] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter. Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery. *IEEE Signal Process. Mag.*, 37(1):105–116, 2020.

[4] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magn. Reson. Med.*, 81(1):439–453, Jan. 2019.

[5] M. Akçakaya, B. Yaman, H. Chung, and J. C. Ye. Unsupervised deep learning methods for biological image reconstruction and enhancement: An overview from a signal processing perspective. *IEEE Signal Process. Mag.*, 39(2):28–44, 2022.

[6] M. Akçakaya, M. Doneva, and C. Prieto (eds.). *Magnetic resonance image reconstruction: theory, methods, and applications*, volume 7 of *Advances in Magnetic Resonance Technology and Applications*. Academic Press, 2022.

[7] Y. U. Alçalar and M. Akçakaya. Zero-shot adaptation for approximate posterior sampling of diffusion models in inverse problems. In *Proc. Eur. Conf. Comput. Vis.*, pages 444–460, 2024.

[8] Y. U. Alçalar and M. Akçakaya. Sparsity-driven parallel imaging consistency for improved self-supervised MRI reconstruction. In *Proc. IEEE Int. Conf. Image Process.*, pages 851–856, 2025.

[9] Y. U. Alçalar, M. Gülle, and M. Akçakaya. A convex compressibility-inspired unsupervised loss function for physics-driven deep learning reconstruction. In *Proc. IEEE Int. Symp. Biomed. Imag.*, pages 1–5, 2024.

[10] Y. U. Alçalar, J. Yun, and M. Akçakaya. Automated tuning for diffusion inverse problem solvers without generative prior retraining. In *Proc. IEEE Int. Workshop CAMSAP*, 2025.

[11] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc. Natl. Acad. Sci.*, 117(48):30088–30095, 2020.

[12] M. Arvinte, S. Vishwanath, A. H. Tewfik, and J. I. Tamir. Deep J-Sense: Accelerated MRI reconstruction via unrolled alternating optimization. In *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, pages 350–360, 2021.

[13] R. Barbano, A. Denker, H. Chung, T. H. Roh, S. Arridge, P. Maass, B. Jin, and J. C. Ye. Steerable conditional diffusion for out-of-distribution adaptation in medical image reconstruction. *IEEE Trans. Med. Imag.*, 44(5):2093–2104, 2025.

[14] E. Bartsch, S. Shin, K. Sheehan, M. Fralick, A. Verma, F. Razak, and L. Lapointe-Shaw. Advanced imaging use and delays among inpatients with psychiatric comorbidity. *Brain Behav.*, 14(2), 2024. Art. no. e3425.

[15] T. A. Basha, M. Akcakaya, C. Liew, C. W. Tsao, F. N. Delling, G. Addae, L. Ngo, W. J. Manning, and R. Nezafat. Clinical performance of high-resolution late gadolinium enhancement imaging with compressed sensing. *J. Mag. Res. Imag.*, 46(6):1829–1838, 2017.

[16] Y. Blau and T. Michaeli. The perception-distortion tradeoff. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 6228–6237, 2018.

[17] B. T. Burdorf. Comparing magnetic resonance imaging and computed tomography machine accessibility among urban and rural county hospitals. *J. Public Health Res.*, 11(1), 2022.

[18] E. J. Candès, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, 2008.

[19] E. P. Chandler, S. Shoushtari, J. Liu, M. S. Asif, and U. S. Kamilov. Overcoming distribution shifts in plug-and-play methods with test-time training. In *Proc. IEEE Int. Workshop CAMSAP*, pages 186–190, 2023.

[20] D. Chen, J. Tachella, and M. E. Davies. Equivariant imaging: Learning beyond the range space. In *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pages 4379–4388, 2021.

[21] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye. Diffusion posterior sampling for general noisy inverse problems. In *Proc. Int. Conf. Learn. Represent.*, 2023.

[22] H. Chung, S. Lee, and J. C. Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *Proc. Int. Conf. Learn. Represent.*, 2024.

[23] H. Chung and J. C. Ye. Score-based diffusion models for accelerated MRI. *Med. Image Anal.*, 80, 2022. Art. no. 102479.

[24] F. Cotter. *Uses of complex wavelets in deep convolutional neural networks*. PhD thesis, University of Cambridge, 2020.

[25] S. U. H. Dar, M. Özbey, A. B. Çatlı, and T. Çukur. A transfer-learning approach for accelerated MRI using deep neural networks. *Magn. Reson. Med.*, 84(2):663–685, 2020.

[26] M. Z. Darestani, A. S. Chaudhari, and R. Heckel. Measuring robustness in deep learning based compressive sensing. In *Proc. Int. Conf. Mach. Learn.*, pages 2433–2444, 2021.

[27] M. Z. Darestani and R. Heckel. Accelerated MRI with un-trained neural networks. *IEEE Trans. Comput. Imag.*, 7:724–733, 2021.

[28] Ö. B. Demirel, S. Moeller, L. Vizioli, B. Yaman, L. Dowdle, E. Yacoub, K. Uğurbil, and M. Akçakaya. High-quality 0.5 mm isotropic fMRI: Random matrix theory meets physics-driven deep learning. In *Proc. Int. IEEE/EMBS Conf. Neural Eng. (NER)*, pages 1–6, 2023.

[29] Ö. B. Demirel, B. Yaman, L. Dowdle, S. Moeller, L. Vizioli, E. Yacoub, J. Strupp, C. A. Olman, K. Uğurbil, and M. Akçakaya. 20-fold accelerated 7T fMRI using referenceless self-supervised deep learning reconstruction. In *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pages 3765–3769, 2021.

[30] O. B. Demirel, B. Yaman, C. Shenoy, S. Moeller, S. Weingärtner, and M. Akçakaya. Signal intensity informed multi-coil encoding operator for physics-guided deep learning reconstruction of highly accelerated myocardial perfusion CMR. *Magn. Reson. Med.*, 89(1):308–321, Jan. 2023.

[31] Y. C. Eldar, A. O. Hero III, L. Deng, J. Fessler, J. Kovacevic, H. V. Poor, and S. Young. Challenges and open problems in signal processing: Panel discussion summary from ICASSP 2017 [panel and forum]. *IEEE Signal Process. Mag.*, 34(6):8–23, 2017.

[32] J. A. Fessler. Optimization methods for magnetic resonance image reconstruction. *IEEE Signal Process. Mag.*, 37(1):33–40, 2020.

[33] D. Gilton, G. Ongie, and R. Willett. Deep equilibrium architectures for inverse problems in imaging. *IEEE Trans. Comput. Imag.*, 7:1123–1133, 2021.

[34] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.*, 47(6):1202–1210, 2002.

[35] H. Gu, B. Yaman, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Revisiting ℓ_1-wavelet compressed-sensing MRI in the era of deep learning. *Proc. Natl. Acad. Sci.*, 119(33), 2022. Art. no. e2201062119.

[36] M. Gülle, J. Yun, Y. U. Alçalar, and M. Akçakaya. Consistency models as plug-and-play priors for inverse problems, 2025. arXiv:2509.22736.

[37] J. Hamilton, D. Franson, and N. Seiberlich. Recent advances in parallel imaging for MRI. *Prog. Nucl. Magn. Reson. Spectrosc.*, 101:71–95, 2017.

[38] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.*, 79(6):3055–3071, 2018.

[39] K. Hammernik, T. Küstner, B. Yaman, Z. Huang, D. Rueckert, F. Knoll, and M. Akçakaya. Physics-driven deep learning for computational magnetic resonance imaging: Combining physics and machine learning for improved medical imaging. *IEEE Signal Process. Mag.*, 40(1):98–114, 2023.

[40] C. Han, T. S. Hatsukami, and C. Yuan. A multi-scale method for automatic correction of intensity non-uniformity in MR images. *Magn. Reson. Med.*, 13(3):428–436, 2001.

[41] R. Heckel, M. Jacob, A. Chaudhari, O. Perlman, and E. Shimron. Deep learning for accelerated and robust MRI reconstruction. *Magn. Reson. Mater. Phys. Biol. Med.*, 37(3):335–368, 2024.

[42] B. Hofmann, I. Ø. Brandsaeter, and E. Kjelle. Variations in wait times for imaging services: a register-based study of self-reported wait times for specific examinations in Norway. *BMC Health Serv. Res.*, 23(1):1287, Nov. 2023.

[43] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya. Dense recurrent neural networks for accelerated MRI: history-cognizant unrolling of optimization algorithms. *IEEE J. Sel. Topics Signal Process.*, 14(6):1280–1291, Oct. 2020.

[44] Y. Hu, W. Gan, C. Ying, T. Wang, C. Eldeniz, J. Liu, Y. Chen, H. An, and U. S. Kamilov. SPICER: Self-supervised learning for MRI with automatic coil sensitivity estimation and reconstruction. *Magn. Reson. Med.*, 2024.

[45] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir. Robust compressed sensing MRI with deep generative priors. In *Proc. Adv. Neural Inf. Process. Syst.*, pages 14938–14954, 2021.

[46] P. M. Johnson, G. Jeong, K. Hammernik, J. Schlemper, C. Qin, J. Duan, D. Rueckert, J. Lee, N. Pezzotti, E. De Weerdt, et al. Evaluation of the robustness of learned MR image reconstruction to systematic deviations between training and test data for the models from the fastMRI challenge. In *Proc. Mach. Learn. Med. Image Reconstruction*, pages 25–34, 2021.

[47] F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn. Reson. Med.*, 81(1):116–128, 2019.

[48] F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akçakaya. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE Signal Process. Mag.*, 37(1):128–140, 2020.

[49] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, et al. fastMRI: a publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiol., Artif. Intell*, 2(1), Jan. 2020. Art. no. e190007.

[50] F. Krueger, C. S. Aigner, K. Hammernik, S. Dietrich, M. Lutz, J. Schulz-Menger, T. Schaeffter, and S. Schmitter. Rapid estimation of 2D relative B1+-maps from localizers in the human heart at 7T using deep learning. *Magn. Reson. Med.*, 89(3):1002–1015, 2023.

[51] T. Küstner, K. Hammernik, D. Rueckert, T. Hepp, and S. Gatidis. Predictive uncertainty in deep learning–based MR image reconstruction using deep ensembles: Evaluation on the fastMRI data set. *Magn. Reson. Med.*, 92(1):289–302, 2024.

[52] G. Luo, M. Blumenthal, M. Heide, and M. Uecker. Bayesian MRI reconstruction with joint uncertainty estimation using diffusion models. *Magn. Reson. Med.*, 90(1):295–311, 2023.

[53] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58(6):1182–1195, Dec. 2007.

[54] J. P. Marques, T. Kober, G. Krueger, W. van der Zwaag, P.-F. Van de Moortele, and R. Gruetter. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage*, 49(2):1271–1281, 2010.

[55] M. Martella, J. Lenzi, and M. M. Gianino. Diagnostic technology: Trends of use and availability in a 10-year period (2011-2020) among sixteen OECD countries. *Healthcare (Basel)*, 11(14), Jul. 2023.

[56] G. McGibney, M. R. Smith, S. T. Nichols, and A. Crawley. Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI. *Magn. Reson. Med.*, 30(1):51–59, 1993.

[57] C. Millard and M. Chiew. A theoretical framework for self-supervised MR image reconstruction using sub-sampling via variable density Noisier2Noise. *IEEE Trans. Comput. Imag.*, 9:707–720, 2023.

[58] V. Monga, Y. Li, and Y. C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.*, 38(2):18–44, 2021.

[59] M. J. Muckley, B. Riemenschneider, A. Radmanesh, S. Kim, G. Jeong, J. Ko, Y. Jun, H. Shin, D. Hwang, M. Mostapha, et al. Results of the 2020 fastMRI challenge for machine learning MR image reconstruction. *IEEE Trans. Med. Imag.*, 40(9):2306–2317, 2021.

[60] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magn. Reson. Med.*, 46(4):638–651, 2001.

[61] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.*, 42(5):952–962, 1999.

[62] P. S. Rajiah, C. J. François, and T. Leiner. Cardiac MRI: State of the art. *Radiology*, 307(3), 2023.

[63] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imag.*, 37(2):491–503, 2018.

[64] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Process. Mag.*, 22(6):123–151, 2005.

[65] E. Shimron, J. I. Tamir, K. Wang, and M. Lustig. Implicit data crimes: Machine learning bias arising from misuse of public data. *Proc. Natl. Acad. Sci.*, 119(13), 2022.

[66] Y. Song, L. Shen, L. Xing, and S. Ermon. Solving inverse problems in medical imaging with score-based generative models. In *Proc. Int. Conf. Learn. Represent.*, 2022.

[67] R. Timofte, E. Agustsson, L. V. Gool, M.-H. Yang, and L. Zhang. NTIRE 2017 challenge on single image super-resolution: Methods and results. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog. Workshop*, pages 114–125, 2017.

[68] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magn. Reson. Med.*, 71(3):990–1001, 2014.

[69] K. Uğurbil, J. Xu, E. J. Auerbach, S. Moeller, A. T. Vu, J. M. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P. F. Van de Moortele, et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage*, 80:80–104, 2013.

[70] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 9446–9454, 2018.

[71] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium, et al. The WU-Minn Human Connectome Project: An overview. *Neuroimage*, 80:62–79, 2013.

[72] L. Winter, J. Periquito, C. Kolbitsch, R. Pellicer-Guridi, R. G. Nunes, M. Häuer, L. Broche, and T. O'Reilly. Open-source magnetic resonance imaging: Improving access, science, and education through global collaboration. *NMR Biomed.*, 37(7), 2024.

[73] B. Yaman, H. Gu, S. A. H. Hosseini, Ö. B. Demirel, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Multi-mask self-supervised learning for physics-guided neural networks in highly accelerated magnetic resonance imaging. *NMR Biomed.*, 35(12), 2022. Art. no. e4798.

[74] B. Yaman, S. A. H. Hosseini, and M. Akcakaya. Zero-shot self-supervised learning for MRI reconstruction. In *Proc. Int. Conf. Learn. Represent.*, 2022.

[75] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magn. Reson. Med.*, 84(6):3172–3191, Dec. 2020.

[76] B. Yaman, C. Shenoy, Z. Deng, S. Moeller, H. El-Rewaidy, R. Nezafat, and M. Akçakaya. Self-supervised physics-guided deep learning reconstruction for high-resolution 3D LGE CMR. In *Proc. IEEE Int. Symp. Biomed. Imag.*, pages 100–104, 2021.

[77] J. Yun, Y. U. Alçalar, and M. Akçakaya. Time-embedded algorithm unrolling for computational MRI, 2025. arXiv:2510.16321.

[78] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, et al. fastMRI: An open dataset and benchmarks for accelerated MRI, 2019. arXiv:1811.08839.

[79] R. Zhao, B. Yaman, Y. Zhang, R. Stewart, A. Dixon, F. Knoll, Z. Huang, Y. W. Lui, M. S. Hansen, and M. P. Lungren. fastMRI+, clinical pathology annotations for knee and brain fully sampled magnetic resonance imaging data. *Scientific Data*, 9(1):152, 2022.

[80] Y. Zhao, Y. Ding, V. Lau, C. Man, S. Su, L. Xiao, A. T. Leong, and E. X. Wu. Whole-body magnetic resonance imaging at 0.05 Tesla. *Science*, 384(6696), 2024.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We confirm that the main claims made in the abstract and introduction accurately reflect our contributions and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have discussed the limitations of out work in extensively, see the Limitations section in Appendix Section G.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: We do not propose a theoretical proof. This question is not applicable to our work.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.

- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We confirm that we have provided a detailed explanation of our experimental setup in Section 4.1 and Appendix Section A. Furthermore, our source codes will be published if the paper is accepted.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Data for our retrospective studies are openly available and details to reproduce the main experimental results are provided in Section 4.1 and Appendix Section A. The code is publicly available at `https://github.com/ualcalar17/CUPID`.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).

- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide a comprehensive explanation of all relevant details in the main text and the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Error bars/standard deviations of quantitative metrics are described in Appendix Section H

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have provided sufficient information on the computer resources in Appendix.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This paper conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We have discussed the potential positive societal impacts of this work in Section 1 and Section 3.1. We believe our work does not pose any negative societal impacts.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We think that our paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The assets including data and baseline models used in this paper are properly cited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: Our paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: We do not do any crowdsourcing or research with human subjects that require instructions

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [Yes]

Justification: The 3D MP2RAGE brain imaging data was acquired on a 7T Siemens Magnetom MRI scanner under the relevant institutional review board (IRB) approvals. Knee and brain imaging data was obtained from the publicly available NYU fastMRI dataset [49, 78], which was collected with IRB approval and subject consent as detailed in the original publication. No additional data involving human subjects were collected in this study.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs were not used in any part of the core methodology, data processing, analysis, or experimental design of this research.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

A Implementation Details for Each Method

PD-DL based approaches For each PD-DL method, VS-QP [2, 75, 29, 30, 76] was unrolled to solve Eq. (3), transforming it into 2 sub-problems:

$$\mathbf{z}^{(i)} = \arg\min_{\mathbf{z}} \|\mathbf{x}^{(i-1)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{10a}$$

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}^{(i)}\|_2^2, \tag{10b}$$

where Eq. (10a) is the proximal operator for the regularization, implicitly solved using neural networks, while Eq. (10b) accounts for data fidelity and has a closed form solution:

$$\mathbf{x}^{(i)} = \left(\mathbf{E}_\Omega^H \mathbf{E}_\Omega + \mu\mathbf{I}\right)^{-1} \left(\mathbf{E}_\Omega^H \mathbf{y}_\Omega + \mu\mathbf{z}^{(i)}\right), \tag{11}$$

which was solved using a conjugate-gradient (CG) method [2] with 15 iterations. Unrolled network for each method comprised 10 unrolls, while the regularizer was implemented as a CNN-based ResNet architecture [67] that had 15 residual blocks. Each layer within these blocks had 3×3 kernels and 64 channels, totaling 592,129 trainable parameters. The unrolled network was trained in an end-to-end fashion for 100 epochs. For supervised PD-DL [38, 2], the normalized ℓ_1-ℓ_2 loss function was used between the reconstructed and ground truth raw k-space data [48, 28]. For SSDU, $\rho = |\Delta|/|\Omega| = 0.4$ was used as proposed in [75]. For EI [20], we modified the loss function in PD-DL networks to:

$$\min_{\boldsymbol{\theta}} \mathbb{E}\left[\mathcal{L}\left(\mathbf{y}_\Omega, \hat{\mathbf{x}}_\Omega\right)\right] + \beta \sum_{g\in G} \mathcal{L}\left(\mathcal{T}_g \hat{\mathbf{x}}_\Omega, f\left(\mathbf{E}_\Omega \mathcal{T}_g \hat{\mathbf{x}}_\Omega, \mathbf{E}_\Omega; \boldsymbol{\theta}\right)\right), \tag{12}$$

where $\hat{\mathbf{x}}_\Omega = f\left(\mathbf{y}_\Omega, \mathbf{E}_\Omega; \boldsymbol{\theta}\right)$ is the PD-DL network output. First term in Eq. (12) enforces consistency with acquired raw data, while the second term imposes equivariance relative to a group of transformations, $\{\mathcal{T}_g\}_{g\in G}$. Here, $|G|$ is the cardinality of $\{\mathcal{T}_g\}_{g\in G}$ and β is the equivariance weight. We followed the authors' publicly available CT reconstruction code for EI [20], and employed 3 rotations along with 2 flips. For CUPID, dual-tree complex wavelet transform (DTCWT), which provides an over-complete representation [64, 24], was selected as the sparsifying transform ($\mathbf{W}$) in Eq. (6). Furthermore, $\mathbf{x}^{(0)}$ in Eq. (6), i.e. the initial estimate prior to any reweighting, was calculated using a CS approach as mentioned in Section 3. This was implemented using Eq. (13) with 10 iterations, with 10 CG steps for data fidelity and $0.01 \cdot ||\mathbf{W}\mathbf{x}_{\text{PI}}||_\infty$ as the soft thresholding parameter.

Compressed sensing We solved the regularized ℓ_1 minimization problem given below:

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega \mathbf{x}\|_2 + \tau\|\mathbf{W}\mathbf{x}\|_1, \tag{13}$$

using VS-QP [32]. Similar to the unrolled network, data fidelity was solved using CG, and soft thresholding was implemented on the DTCWT coefficients.

DDS We followed the original code and pre-trained score network provided by [22] in their corresponding public repository. We used 50-150 DDIM sampling steps depending on the SNR and noise level, and tuned the number of CG iterations for the best performance.

ScoreMRI For ScoreMRI implementation, we followed the original code and pre-trained network provided by [23] in their corresponding public repository and used 2000 predictor-corrector (PC) sampling.

Each method (including CUPID) was implemented using a single NVIDIA A100-SXM4-80GB GPU that has a total memory of 80 GB.

B Computational Efficiency and Runtime Analysis

As noted in Section 4.2 and Appendix A, all PD-DL comparison methods share the same unrolled network architecture. Consequently, they exhibit identical inference times of under one second per slice once they are trained on a database. Diffusion-based methods, such as ScoreMRI and DDS, require substantially longer training and sampling times due to their iterative denoising processes and

Table 2: Training and inference/run-time of comparison methods under database and zero-shot settings using a single NVIDIA A100-SXM4-80GB GPU. N/A indicates methods without dataset-level training or zero-shot configuration.

Method	Database setting		Zero-shot setting
	Train [GPU hours] ↓	Inference ↓	Runtime ↓
Supervised [38]	∼3.5 hours	$\ll$ 1 seconds	-
SSDU [75]	∼3.5 hours	$\ll$ 1 seconds	-
EI [20]	∼3.5 hours	$\ll$ 1 seconds	-
ScoreMRI [23]	∼168 hours	∼350 seconds	-
DDS (150 NFEs) [22]	∼168 hours	∼10 seconds	-
ZS-SSDU [75]	-	-	$\sim \mathrm{K}_{\mathrm{ZS-SSDU}}\times$15 seconds
CUPID **(ours)**	$\sim (\mathrm{K}_{\mathrm{CUPID}}+1)\times$3.5 hours	$\ll$ 1 seconds	$\sim (\mathrm{K}_{\mathrm{CUPID}}+1)\times$63 seconds

large model complexity. These methods often involve hundreds to thousands of sampling steps during inference, resulting in inference times that are orders of magnitude higher than PD-DL approaches.

In the zero-shot setting, ZS-SSDU serves as a useful baseline that, like CUPID, adapts to individual scans without dataset-level retraining. However, ZS-SSDU relies on access to raw k-space data and employs $\mathrm{K}_{\mathrm{ZS-SSDU}} = 25$ k-space masks in its public implementation. In contrast, CUPID operates entirely on reconstructed DICOM images and eliminates the need for raw data while maintaining comparable zero-shot reconstruction performance.

Although zero-shot optimization introduces additional computation, such run-times are generally acceptable in clinical workflows where reconstructions can be completed offline, for instance when image readings occur on the following day [15]. Furthermore, CUPID introduces a controllable trade-off between reconstruction quality and computational cost, determined by the number of perturbation patterns used during optimization (denoted as $\mathrm{K}_{\mathrm{CUPID}}$ in Tab. 2). In our experiments, we set $\mathrm{K}_{\mathrm{CUPID}} = 6$, resulting in a runtime of approximately 7–8 minutes per slice, depending on the network depth. Both factors can be adjusted to balance reconstruction fidelity against run-time.

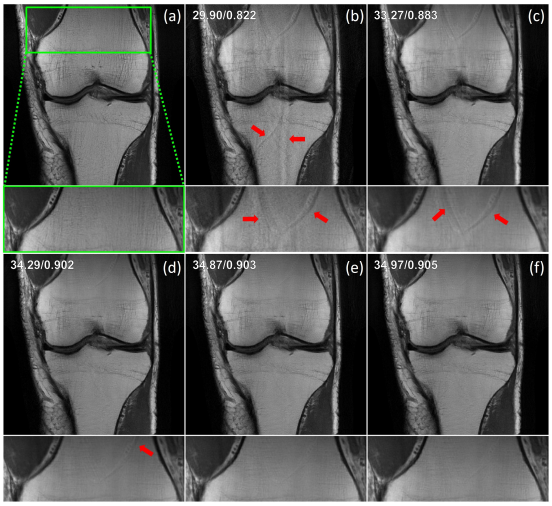
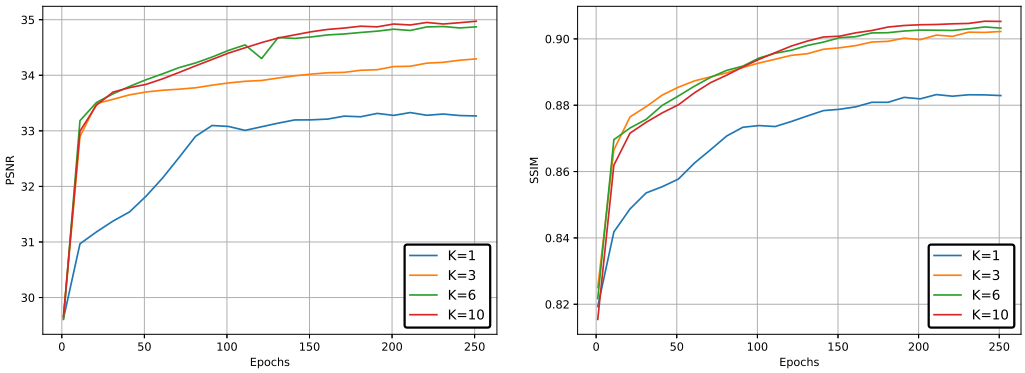


Figure 8: Representative zero-shot fine-tuning result for CUPID with varying perturbation counts (K) on coronal PD knee MRI ($R = 4$). (a) Fully-sampled, (b) CG-SENSE, (c–f) CUPID with $K = 1, 3, 6, 10$. Fewer perturbations ($K < 6$) yield artifacts due to poor expectation estimates; quality improves with K, but gains become negligible beyond $K = 6$.

C Perturbation Strategies

C.1 Choice for number of perturbation patterns

The empirical expectation that approximates the one in Eq. (7) is expected to converge to the true expectation as we introduce more perturbation patterns and randomness over the choice of $\mathbf{p}$. Fig. 8 shows the zero-shot fine-tuning results of CUPID with $K \in \{1, 3, 6, 10\}$, while Fig. 9a and Fig. 9b illustrates the corresponding PSNR and SSIM curves throughout the training epochs, respectively. As expected, using a single pattern does not capture the true mean and exhibits artifacts. As we introduce more perturbations, we reduce the artifacts and noise amplification. At a certain point, increasing the number of perturbations becomes counterproductive, yielding only marginal gains while significantly increasing the computation time. Thus, we opted to use 6 distinct $\mathbf{p}_{\mathrm{k}}$ patterns throughout our study as it offers the optimal trade-off.



(a) PSNR curves for each K value.

(b) SSIM curves for each K value.

Figure 9: PSNR and SSIM curves confirm the visual observations with respect to the number of perturbations, K. Lower K values tend to perform worse and increasing K becomes redundant after a certain point.

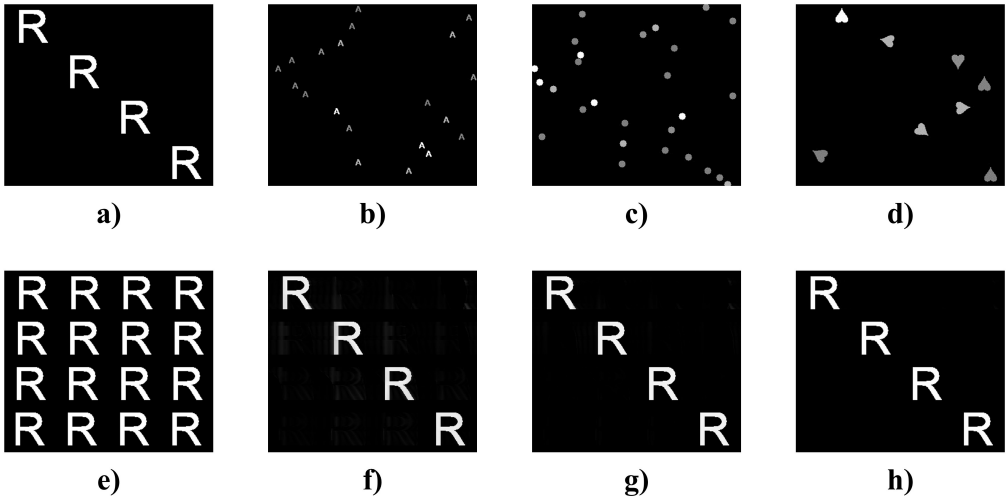


Figure 10: Added perturbations may consist of: (a) precisely positioned letters that have the same intensity in within each shape, (b) randomly positioned letters or (c) circles that have different intensities, or (d) randomly rotated card suits. Furthermore, when accelerated by $R = 4$, even without ACS data, whose corresponding zerofilled image is shown in (e), these perturbations can be resolved via parallel imaging methods such as CG-SENSE: (f) 20 iterations, (g) 40 iterations, (h) 80 iterations.

C.2 Design alternatives for perturbations

As stated in Section 3, added perturbations may consist of several different structures. Fig. 10 provides some of these perturbation examples, an illustration of how the perturbation looks with undersampling, and how they are recovered perfectly through conventional parallel imaging methods.

We note that there was no task-specific perturbation that we used, meaning that the perturbations selected from the same set were applied to all datasets given that the created perturbations do not create fold-overs at R-fold which result in artifacts. Note the latter condition means they should be recoverable through parallel imaging reconstruction. Finally, we note that when calculating the sample mean estimate for Eq. (7), intensity of the perturbations was empirically found to be more important than their shapes/orientations. Specifically, we observed that varying it randomly within the perturbation, as in Fig. 10b-d, leads to improved reconstruction outcomes.

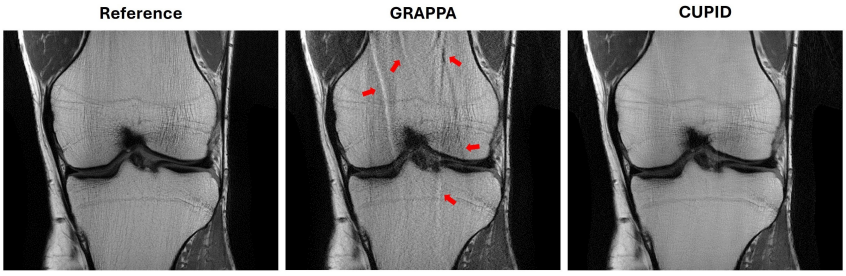


Figure 11: Representative reconstructions for CUPID with $\mathbf{x}_{\text{PI}}$ reconstructed using GRAPPA on coronal PD knee MRI using $R = 4$ uniform undersampling. GRAPPA exhibits aliasing and noise artifacts at this high acceleration rate. PD-DL network trained with a CUPID implementation that only has access to this GRAPPA reconstruction improves on it, reducing these artifacts. This highlights the compatibility of CUPID with different parallel imaging reconstructions.

D Compatibility with Various Parallel Imaging Reconstructions

Vendor reconstructions typically use different parallel imaging techniques. For our retrospective studies, we used CG-SENSE (or equivalently SENSE) [61] because it naturally fits with the DF units in the unrolled network, and it is commonly used in clinical settings, alongside GRAPPA [34]. However, we emphasize that our method does not make assumptions about the specific reconstruction method used by the vendor; instead, it assumes that parallel imaging can resolve the perturbations, which is ensured by designing them in a manner that prevents fold-over aliasing artifacts from overlapping.

To further validate this, we include representative CUPID reconstruction results in Fig. 11 where $\mathbf{x}_{\text{PI}}$ is generated via GRAPPA [34], demonstrating that CUPID is compatible with different types of parallel imaging reconstructions as input. We further note that the prospective study also used GRAPPA reconstruction as input, as this is the reconstruction provided by the vendor used in our institution.

E Effect of the Trade-off Parameter

We explored the effect of λ parameter in CUPID by training 5 distinct PD-DL networks using $\lambda \in \{0, 50, 100, 200, \infty\}$. We note that using $\lambda = 0$ corresponds to using only the compressibility term ($\mathcal{L}_{\text{comp}}$ in Eq. (6)), whereas using $\lambda \to \infty$ translates to using solely the parallel imaging fidelity term ($\mathcal{L}_{\text{pif}}$ in Eq. (7)). Fig. 12 shows the corresponding reconstruction results for each case. As outlined in Section 3, only using $\mathcal{L}_{\text{comp}}$ leads to overly-smooth reconstructions due to network forcing the wavelet coefficients towards zero without maintaining consistency with the data. On the other hand, solely using $\mathcal{L}_{\text{pif}}$ results in DIP-like reconstructions [70], where the network overfits the data without any regularization, resulting in noise amplification. CUPID with $\lambda \in \{50, 100, 200\}$ combines both loss terms to attain high-fidelity reconstructions. Thus, we conclude that CUPID

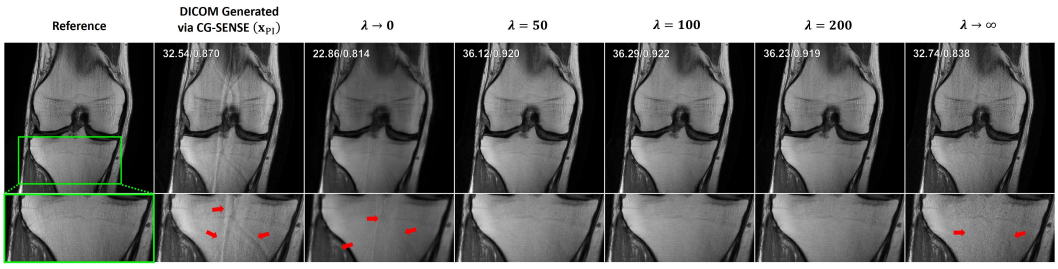


Figure 12: Using only the compressibility term ($\lambda = 0$) in the loss leads to overly-smoothed images, whereas using only the parallel imaging fidelity term ($\lambda \to \infty$) causes noise amplification (red arrows). Using both terms with a mid-range λ value as a trade-off provides high-quality reconstructions that are clear from noise and artifacts.

demonstrates robust performance across a wide range of λ values, provided that λ is chosen within a reasonable range.

F Experiments on Sampling Pattern Variations and High Acceleration Rates

We further evaluated the robustness of CUPID to variations in acceleration rate and sampling pattern using the fastMRI knee and brain datasets [49, 78]. Representative zero-shot reconstruction results for these settings are shown in Fig. 13.

Specifically, we tested:

- Random uniform undersampling at an acceleration rate of $R = 4$ (in contrast to the equidistant pattern used in the main text),
- Equidistant undersampling at higher acceleration rates of $R = 6$ and $R = 8$.

For random uniform undersampling at $R = 4$, the results were comparable to those obtained with equidistant sampling, exhibiting similarly clean reconstructions with effective noise suppression and minimal artifacts. At $R = 6$, CUPID continued to produce high-quality reconstructions, preserving image sharpness and structure with only minor residual artifacts.

At $R = 8$, CUPID still achieved effective noise suppression but exhibited visible residual artifacts in some regions, indicating that this acceleration factor likely represents the practical limit with the current coil configurations and SNRs on the *fastMRI* datasets. The observed degradations arise primarily from the increasing loss of quality for the parallel imaging reconstruction at high acceleration, which weakens the initialization, and from the general limitations of DL-based methods in highly underdetermined regimes. Previous studies have shown that supervised PD-DL methods yield visible blurring and reduced diagnostic quality at $R = 8$ [59], while self-supervised approaches trained with raw k-space data also exhibit residual aliasing at this rate [75, 8]. Since CUPID operates without any access to raw k-space measurements, similar or slightly greater degradation at extreme acceleration rates is expected. Overall, these results demonstrate that CUPID generalizes effectively across sampling patterns and maintains strong reconstruction performance up to moderate acceleration factors, with performance degradation at very high acceleration factors consistent with known challenges in highly accelerated MRI reconstruction.

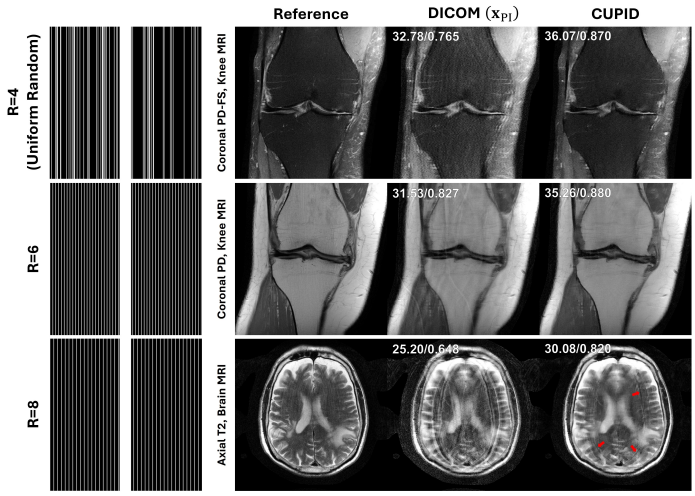


Figure 13: Representative reconstructions from the fastMRI dataset for different acceleration rates and sampling patterns, illustrating robustness up to $R = 6$. Artifact emerge at $R = 8$, as expected.

G Limitations

We note some practical aspects for the translation of CUPID. While our method aims to improve equitable access to fast MRI in low-resource settings, we acknowledge hardware limitations for fine-tuning PD-DL models; however, since CUPID only requires anonymized DICOM images, data can be securely transferred to off-site locations with greater computational resources. In addition, when a reference calibration scan is unavailable, as in most DICOM databases, coil sensitivities must be estimated during reconstruction. This is addressed in standard PD-DL with raw k-space access [44, 12], but outside the scope of this study.

MR scans may also include filtering operations applied by some vendors that affect the assumption $\hat{\mathbf{x}}_{\text{PI}} = (\mathbf{E}_\Omega^H \mathbf{E}_\Omega)^{-1} \mathbf{E}_\Omega^H \mathbf{y}_\Omega$. This was discussed extensively in [65], in the context of using retrospective undersampling of DICOM images to train DL reconstruction, especially highlighting the use of zero-padding, which improves the display resolution compared to the acquisition resolution. It was shown that training of models from retrospective undersampling of DICOM image databases for PD-DL training using zero-padding may lead to biases and inaccuracies. Conversely, our approach is physics-driven in nature, and the sampling pattern Ω naturally accounts for the zero-padding operation. However, our method is not immune to other types of filtering/processing, such as implicit intensity correction [40] or deidentification methods [71], in which case the filtered $\mathbf{x}_{\text{PI}}$ would need to be treated as the parallel imaging solution corresponding to a filtered version of $\mathbf{y}$.

H Additional Quantitative Analysis with Standart Deviation

To provide a more comprehensive comparison, Tab. 3 reports the mean and standard deviation for all quantitative metrics across four datasets. This supplements Tab. 1 by quantifying the consistency of each method.

Table 3: Quantitative results with standard deviations for all comparison methods on Coronal PD, Coronal PD-FS, Ax FLAIR, and Ax T2 datasets using equispaced $R = 4$ undersampling. Mean $\pm$ std values are shown. Categories and highlight conventions follow Tab. 1 in the main text.

Method	Cor PD, Knee MRI		Cor PD-FS, Knee MRI		Ax FLAIR, Brain MRI		Ax T2, Brain MRI	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Supervised [38]	40.44 ± 3.27	0.964 ± 0.018	35.72 ± 2.49	0.893 ± 0.058	37.91 ± 1.81	0.967 ± 0.008	36.77 ± 3.25	0.933 ± 0.062
SSDU [75]	39.64 ± 3.26	0.957 ± 0.021	35.68 ± 2.49	0.892 ± 0.060	37.55 ± 1.76	0.964 ± 0.009	36.59 ± 3.11	0.931 ± 0.060
EI [20]	38.07 ± 3.79	0.938 ± 0.033	33.83 ± 2.82	0.849 ± 0.075	36.40 ± 1.70	0.953 ± 0.010	34.58 ± 2.88	0.909 ± 0.066
ScoreMRI [23]	37.85 ± 3.66	0.928 ± 0.031	35.01 ± 2.46	**0.883 ± 0.058**	35.26 ± 1.84	0.934 ± 0.013	33.83 ± 2.70	0.899 ± 0.065
DDS [22]	38.64 ± 3.42	0.950 ± 0.026	34.89 ± 2.50	0.875 ± 0.064	36.18 ± 1.95	0.950 ± 0.012	35.16 ± 3.13	0.914 ± 0.076
PI [60, 34]	35.51 ± 4.14	0.909 ± 0.047	29.48 ± 3.43	0.735 ± 0.112	31.97 ± 1.99	0.906 ± 0.018	31.02 ± 2.49	0.833 ± 0.075
CS [53]	36.92 ± 3.72	0.922 ± 0.035	33.31 ± 2.71	0.841 ± 0.079	33.17 ± 1.79	0.929 ± 0.013	33.61 ± 2.61	0.897 ± 0.071
CUPID **(ours)**	**38.82 ± 3.23**	**0.952 ± 0.024**	**35.04 ± 2.45**	0.880 ± 0.059	**36.49 ± 1.72**	**0.957 ± 0.009**	**35.31 ± 3.09**	**0.921 ± 0.073**

I More Results from the Retrospective Study

We provide additional qualitative reconstruction results for various methods, showcasing representative reconstructions from each dataset, as illustrated in Fig. 14.

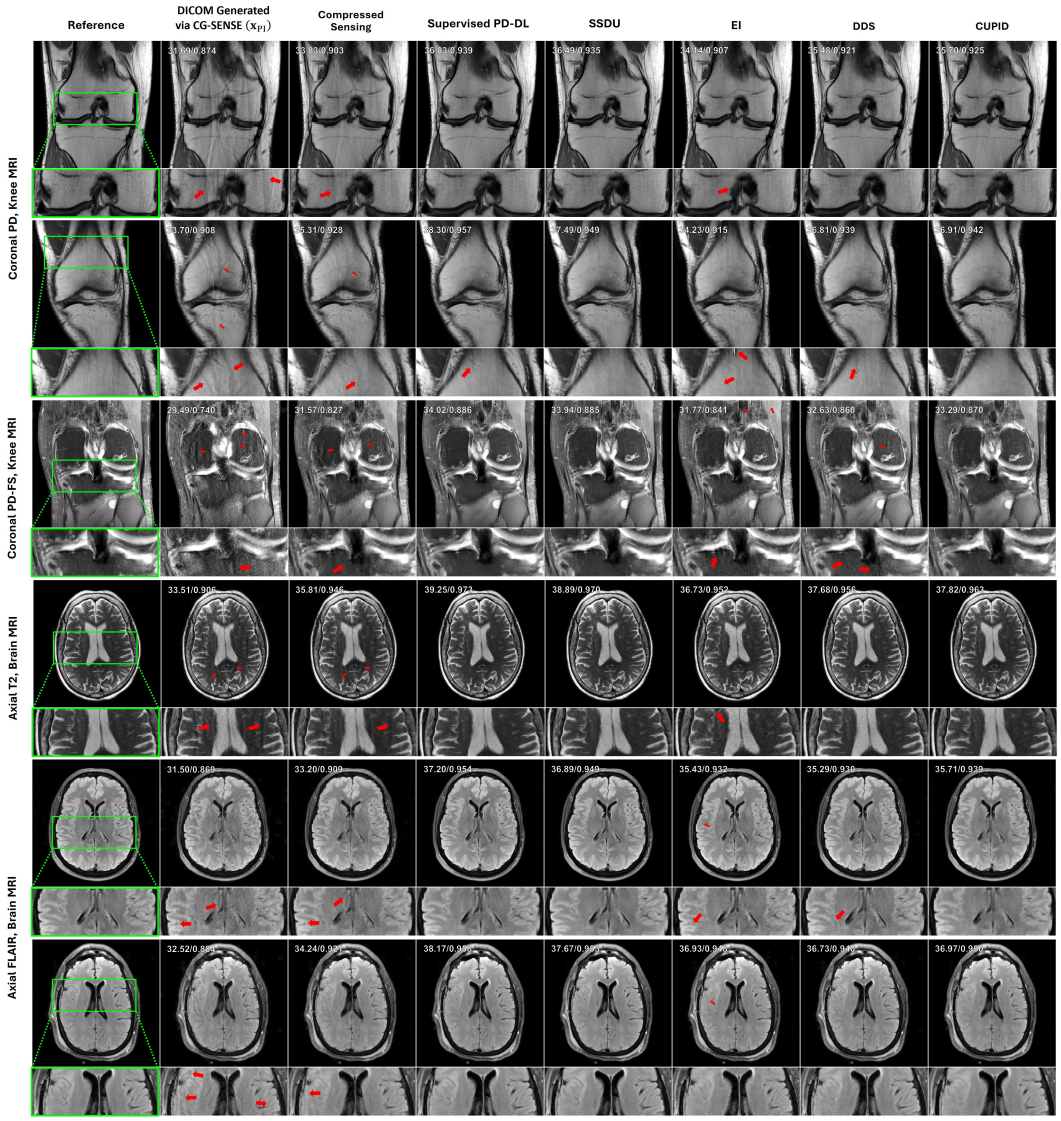


Figure 14: Illustrative slices reconstructed at $R = 4$ using equidistant undersampling from coronal PD and coronal PD-FS knee MRI, as well as axial T2 and FLAIR brain MRI. CG-SENSE, CS, EI and DDS exhibit artifacts. On the other hand, CUPID surpasses them with high-fidelity reconstructions, closely matching supervised and self-supervised PD-DL methods that require raw k-space data during training.