Architecture-Agnostic Untrained Network Priors for Image Reconstruction with Frequency Regularization

Yilin Liu[1], Yunkui Pang[1], Jiang Li[1], Yong Chen[2], and Pew-Thian Yap[1]

[1] Computer Science, University of North Carolina at Chapel Hill
[2] Radiology, Case Western Reserve University
{yilinliu,yunkuipa,jianglz}@cs.unc.edu, yxc235@case.edu, ptyap@med.unc.edu

Abstract. Untrained networks inspired by deep image priors have shown promising capabilities in recovering high-quality images from noisy or partial measurements *without requiring training sets*. Their success is widely attributed to implicit regularization due to the spectral bias of suitable network architectures. However, the application of such network-based priors often entails superfluous architectural decisions, risks of overfitting, and lengthy optimization processes, all of which hinder their practicality. To address these challenges, we propose efficient architecture-agnostic techniques to directly modulate the spectral bias of network priors: 1) bandwidth-constrained input, 2) bandwidth-controllable upsamplers, and 3) Lipschitz-regularized convolutional layers. We show that, with *just a few lines of code*, we can reduce overfitting in underperforming architectures and close performance gaps with high-performing counterparts, minimizing the need for extensive architecture tuning. This makes it possible to employ a more *compact* model to achieve performance similar or superior to larger models while reducing runtime. Demonstrated on inpainting-like MRI reconstruction task, our results signify for the first time that architectural biases, overfitting, and runtime issues of untrained network priors can be simultaneously addressed without architectural modifications. Our code is publicly available [3].

1 Introduction

Magnetic resonance imaging (MRI) is a mainstream imaging tool for medical diagnosis. Reconstructing MR images from raw measurements involves data transformation from a Fourier spectrum in k-space to image space [11, 28]. Since acquiring full k-space measurements is time-consuming, under-sampled k-space data are often collected to reduce scan times. This makes accelerated MRI reconstruction an ill-posed inverse problem that conventionally requires handcrafted priors [25, 30] to mitigate the resulting aliasing artifacts. While supervised learning methods based on convolutional neural networks (CNNs)

[3] https://github.com/YilinLiu97/Untrained-Recon.git

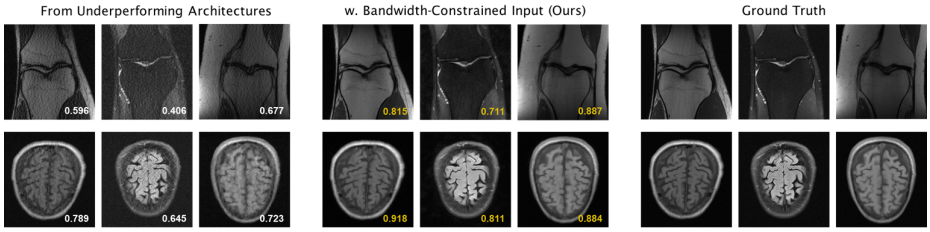


Fig. 1: Improving underperforming architectures (SSIM (↑)). Turning the left to the right simply by replacing the white-noise input with *selected* Fourier features or low-pass filtering the noise input, which can be implemented with *a few lines of code*.

have enhanced reconstruction quality with fewer measurements, their training relies on paired under-sampled and fully-sampled measurements, which are expensive to acquire and can affect robustness and generalization across different acquisition protocols or anatomical variations [21, 22].

Instead of requiring large-scale datasets for capturing prior statistics, untrained networks [12] inspired by deep image prior (DIP) [43] use only the corrupted or partial measurements and a task-specific forward operator. Reconstruction is regularized solely by the inductive biases of the network architectures, enabling zero-shot self-supervised reconstruction for various imaging inverse problems [26, 36, 48]. Concretely, a CNN, which parameterizes the unknown desired image, is optimized such that the output image, transformed by the forward operator, matches the acquired measurements. Such parameterization exhibits surprisingly high resistance to noise and corruption, which acts as a form of implicit regularization. Recent studies have attributed this property to CNN's inherent spectral bias—the tendency to fit the low-frequency signals before the high-frequency signals (e.g., noise) [4, 39]. The choice of network architecture is shown to be critically relevant to such bias [1, 4, 29].

Despite the great promise, obtaining favorable results with untrained network priors is contingent upon two critical factors: an optimal architecture specific to a task and an early-stopping strategy to prevent overfitting to noisy or partial measurements. Furthermore, optimizing on a per-image basis makes the reconstruction process domain-agnostic but inherently slow [46]. While these issues are intertwined, with overfitting and runtime issues exacerbated by inappropriate and over-parameterized architectures, most existing efforts tackle these challenges separately. For architectural design, existing methods rely on either handcrafting or utilizing neural architecture search techniques [1, 5, 6, 12, 14, 24]. However, the lack of consensus on architectural priors often leads to laborious search. Another line of work is dedicated to preventing overfitting through oracle early-stopping criterions [44, 46], subspace optimization [2] or pretraining then fine-tuning [3, 34]. These methods mostly use fewer trainable parameters or hold out a subset of measurements for self-validation, in the spirit of traditional strategies, and often trade-off accuracy or involve costly pre-training.

In this work, we explore the possibility of modulating the frequency bias and hence the regularization effects of network priors in an *architecture-agnostic manner*, aiming to enhance the performance of a given architecture irrespective of its configuration specifics. This is conceivable in light of the recent body of theoretical and empirical evidence indicating that there are *only a few* key components (e.g., unlearnt upsampling) within the architecture that are the driving forces behind the spectral bias in DIP [4, 13, 29, 39]. Motivated by these findings, we develop efficient methods from a frequency perspective to effectively regularize the network priors, alleviating overfitting by curbing the overly rapid convergence of undesired high-frequency components, all with minimal architectural modifications and computational costs. Specifically, we propose to (1) constrain the effective bandwidth of the input via blurring or using Fourier features with a *narrower* frequency range, (2) adjust the bandwidths of the interpolation-based upsamplers with controllable attenuation (smoothing) extents, and (3) regularize the Lipschitz constants of the convolutional layers to enforce function smoothness.

We found empirically that by mitigating convergence to high-frequency components, our regularized network priors not only exhibit less vulnerability to overfitting but also tend to achieve better extrapolation capabilities in inpainting tasks. In the context of MRI reconstruction, which is essentially an inpainting task occurring in k-space, our methods significantly improve models across various architectural configurations without necessitating extensive architectural tuning (Fig. 1). Their efficacy is also showcased in denoising and inpainting for natural images. By minimizing architectural influences, our approach additionally offers a unique advantage in efficiency: a **smaller**, previously underperforming network, can now achieve performance on-par with or even surpasses a **larger**, heavily parameterized high-performing network. Our contribution is three-fold:

- We propose efficient techniques to directly modulate the frequency bias in untrained network priors, addressing architectural design, overfitting, and runtime challenges in a unified, architecture-agnostic manner.
- The enhanced untrained networks match leading self-supervised methods with up to 90× faster runtime (1 hr/slice to ∼5 min/slice) for MRI reconstruction and surpass supervised methods on out-of-domain data.
- Our findings on medical and natural image reconstruction reveal the spectral behaviors of CNNs in a single-instance generative setting.

2 Related Work

Spectral Bias and function smoothness. Function smoothness, also referred to as function frequency, quantifies how much the output of a function varies with changes in its input [9]. Spectral bias [37,45] is an implicit bias that favors learning functions changing at a slow rate (low-frequency), e.g., functions with a small Lipschitz constant. In visual domains, this is evident from the lack of subtle details in network outputs. Many regularization techniques to aid generalization, such as early stopping and ℓ_2 regularization [33, 38], implicitly encourage smoothness. To explicitly promote smoothness, it is natural to penalize the norm of the

input-output Jacobian [15, 35]. However, computation of the Jacobian matrix for the high-dimensional MRI data during training is very expensive. Another efficient and prevalent solution is to constrain the network to be c-Lipschitz with a *pre-defined* Lipschitz constant c [10, 32]. We followed this line with a novel aim of achieving architecture-agnostic untrained image reconstruction.

Input frequency and overfitting. The network input plays an important role in helping the neural network represent signals of various frequencies. Pioneering work on spectral bias [37] showed theoretically and empirically that fitting high frequencies becomes easier, provided that the data manifold itself contains high-frequency components (Sec. 4 in [37]). This has directly motivated implicit neural representations (INRs) [40] and neural radiance fields (NeRFs) [31] where coordinates are mapped to RGB values: naively training with raw coordinates as inputs results in over-smoothing; encoding the input coordinates with sinusoidal functions of higher frequencies enables the network to represent higher frequencies [31, 41]. However, it has recently been reported that the high-frequency bands of NeRF's input encodings incur overfitting and lead to failure in few-shot settings [47]. We found similar issues in untrained networks and propose to constrain the input's frequency range to counteract the convergence of high frequencies.

Architecture-induced spectral bias. Recent studies on the working mechanisms of DIP reveal that unlearnt upsampling, with the low-pass characteristics of its interpolation filter, is responsible for the regularizing effects of DIP [4, 13, 29]. Liu et al. [29] showed that the fixed upsampling operations readily bias the architecture towards producing low-frequency outputs, critically influencing both the peak PSNR and the starting point at which performance decays. The convolutional layer is another core element that exhibits stronger frequency selectivity compared to fully-connected layers and 1D layers [4, 39]. These findings motivated us to operate directly on these elements to achieve architecture-agnostic control.

Avoiding overfitting is a primary goal of unsupervised reconstruction where only noisy or partial measurements are available. Wang et al. [44] proposed an early-stopping (ES) criterion by tracking the running variance of the output, but it is found to be unstable in medical settings [2]. Yaman et al. [46] proposed to split the available measurements into training and validation subsets and use the latter for self-validation for automated early stopping. While ES prevents overfitting, it cannot enhance the capability of a given architecture like our approach. Transfer-learning based methods aim to use fewer non-trainable parameters by performing pre-training followed by fine-tuning [3, 34] or subspace optimization [2]. In contrast, our method directly modulates the spectral bias to mitigate the convergence of undesired high frequencies, which is found to also improve the networks' extrapolation capabilities, all while maintaining the model complexity.

3 Method

3.1 Preliminaries

Accelerated MRI The goal of accelerated MRI reconstruction is to recover a desired image $\mathbf{x} \in \mathbb{C}^n$ ($n = n_{\mathrm{h}} \times n_{\mathrm{w}}$) from a set of under-sampled k-space

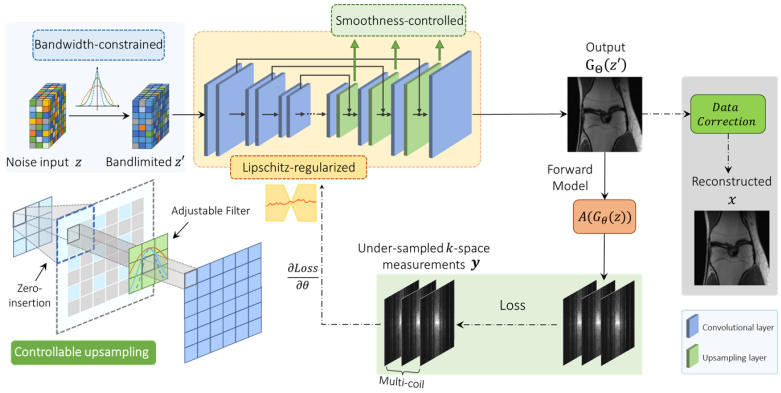


Fig. 2: Overview of the proposed regularized network priors for MRI reconstruction. By adjusting the bandwidth of the input, the interpolation-based upsampling, and regularizing the convolutional layers, our approach enables more direct control over the spectral bias of architectures with various depths and widths.

measurements. We focus on a multi-coil scheme in which the forward model is defined as

$$\mathbf{y}_i = \mathbf{A}_i\mathbf{x} + \epsilon, \quad \mathbf{A}_i = \mathbf{MFS}_i, \quad i = 1, \ldots, c, \tag{1}$$

where $\mathbf{y}_i \in \mathbb{C}^m$ denotes the k-space measurements from coil i, c denotes the number of coils, $\mathbf{S}_i \in \mathbb{C}^n$ denotes the coil sensitivity map (CSM) that is applied to the image $\mathbf{x}$ through element-wise multiplications, $\mathbf{F} \in \mathbb{C}^{n\times n}$ denotes the 2D discrete Fourier transform, $\mathbf{M} \in \mathbb{C}^{m\times n}$ denotes the under-sampling mask, and $\epsilon \in \mathbb{C}^m$ denotes the measurement noise.

Untrained MRI Reconstruction is often framed as an inpainting problem where the network recovers the unacquired k-space measurements (masked) based on the acquired k-space data (observed). The image $\mathbf{x}$ is parameterized via a neural network $\mathbf{G}_\theta(\mathbf{z})$ with a fixed noise input vector $\mathbf{z}$ drawn from a uniform distribution $z \sim \mathcal{U}(0, 1)$. With the MRI forward model in Eq. 1, the untrained network solves the following optimization problem:

$$\theta^* = \arg\min_\theta \mathcal{L}(\mathbf{y}; \mathbf{AG}_\theta(\mathbf{z})), \quad \mathbf{x}^* = \mathbf{G}_{\theta^*}(\mathbf{z}). \tag{2}$$

This parameterization enables the design of novel image priors based on the network architecture and its parameters, rather than on handcrafted image space priors. Nevertheless, many studies augment the untrained networks with traditional image regularizers [27], e.g., total variation (TV), which promotes piecewise constant images and can only partially alleviate over-fitting [2, 34].

3.2 Architecture-Agnostic Frequency Regularization

To modulate the regularization effects of the network prior, we identify three core elements within the framework of DIP that lead to spectral bias and introduce corresponding regulation methods, as depicted in Fig. 2.

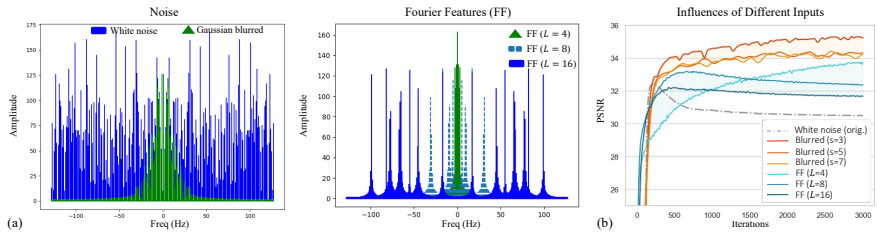


Fig. 3: (a) Visualization of the 1D white noise, the low-passed noise via Gaussian blur, and Fourier features of various frequencies in the frequency domain. (b) Limiting the input's bandwidth via either Gaussian blur or Fourier features with lower f_c ($L = 4$ or 8) are both effective in alleviating overfitting and enhancing the peak performance.

Bandwidth-Constrained Input. Inspired by implicit neural representations (INRs), we rethink the role of inputs in untrained networks for learning different frequencies. Conventionally, the inputs are randomly sampled from either a uniform or Gaussian distribution and are then mapped to image intensities. From a frequency perspective, such input comprises all frequencies with uniform intensities [8], as white noise with variance σ^2 exhibits an autocorrelation that is a scaled Dirac δ-function $\sigma^2\delta(t)$, whose Fourier transform $\mathcal{F}$ has a constant magnitude σ^2 spanning all frequencies μ, i.e., $\mathcal{F}\{\sigma^2\delta(t)\}(\mu) = \sigma^2$, $\mu \in \mathbb{R}$. With this view in mind, we draw an analogy between untrained networks and INRs that map Fourier features to RGB values. Fourier features are sinusoid functions of the input coordinates $\mathbf{p}$, i.e., $[\sin(2^0\pi\mathbf{p}), \cos(2^0\pi\mathbf{p}), ..., \sin(2^{L-1}\pi\mathbf{p}), \cos(2^{L-1}\pi\mathbf{p})]$, where a larger L assists the network in representing higher frequencies [41]. This is consistent with [37] which indicates that the frequency magnitudes that can be expressed by the network increase with increasing frequency in the data manifold.

In this sense, an untrained network can be viewed as *mapping a broad spectrum of Fourier features to a target image* (Fig. 3 (a)). We hypothesize that this promotes rapid convergence of high frequencies, which is likely to bias the network towards high-frequency artifacts and hinder its ability to exploit spatial information effectively for extrapolation.

To validate this, we applied a Gaussian blur filter $\mathcal{G}_{s,\sigma}$ on the noise input z to remove a certain amount of high frequencies before passing it to the network: $z * \mathcal{G}_{s,\sigma}$, where $*$ denotes convolution. The filter size s and sigma value σ that controls the filter's bandwidth are hyperparameters. As exemplified in Fig. 4, simply adjusting σ already brings significant gains without architectural changes. Alternatively, as shown in Fig. 3 (b), substituting the

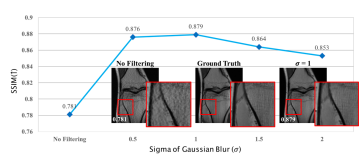


Fig. 4: Narrowing input's bandwidth promotes low frequencies and enhances extrapolation capabilities as a "free lunch". The output becomes smoother as σ increases, up to a certain point.

noise input with Fourier features, with a carefully selected maximum frequency $f_c \propto L$ (e.g., $L = 4$ or 8) to narrow the input's effective bandwidth, is also effective. As L increases, e.g., $L = 16$, the frequency range of Fourier features input approximates that of the original noise input, and the performance deteriorates similarly, further supporting our hypothesis. Fourier features introduce frequency-diverse input akin to white noise but in a *controlled* manner, enabling regularization over the frequency selectivity of the network.

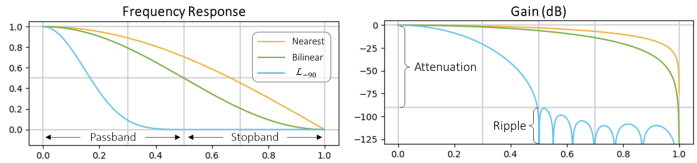
In light of this empirical evidence, we propose to *limit* the input's bandwidth to mitigate the fitting of high-frequency components in untrained networks, achievable efficiently via blurring or Fourier features with a narrowed frequency range. Gaussian blur and Fourier Features offer flexible bandwidth control over the input through hyperparameters $\{s, \sigma\}$ and L, respectively, which allows for scaling under higher noise level/undersampling rates (examples of 8× undersampling in Fig. 8). We experimented with more sophisticated schedulers for the hyperparameters, but found that simply fixing them throughout the training yields superior performance.

Bandwidth-Controllable Upsampling. We observed that only constraining the input's bandwidth significantly enhances *shallower* architectures, yet the improvement diminishes as the network depth increases (Tab. 2, 3). This could be partly attributed to the increased network layers that can always generate new arbitrarily high frequencies [19, 37].

We note that the interpolation-based upsampling methods within the network, such as nearest neighbor and bilinear, essentially act as implicit low-pass filters, attenuating the alias frequencies caused by the increased sampling rate of the input feature maps. Prior works [13, 29] in denoising have shown that these non-trainable upsampling methods are driving forces behind the spectral bias of DIP, delaying the convergence of higher frequencies. Different upsamplers bias the network towards different spectral properties, depending on the bandwidth of the interpolation filter [29]. Here, we show in Tab. 1 that upsampling also substantially influences the performance image reconstruction.

Table 1: Influences of upsamplers on reconstruction. From the left to the right, the attenuation extent of the upsampling method increases. Construction details of the upsampler $\mathcal{L}_{-90}$ follows [29]. Frequency responses of the interpolation filters are shown in the figure below. Evaluated on the 4× fastMRI multi-coil brain datasets.

Methods	w/o. Upsampling.	Nearest	Bilinear	$\mathcal{L}_{-90}$	# of Params. (Millions)
ConvDecoder [6]	28.69 ± 1.6	31.78 ± 1.2	32.31 ± 1.3	32.48 ± 1.2	4.1 M
Deep Decoder [12]	24.55 ± 1.1	27.10 ± 0.9	31.36 ± 1.4	32.68 ± 1.1	0.47 M



Motivated by these results, we introduce an upsampler with *controllable bandwidth* so as to modulate the network's spectral bias, especially for *deeper* architectures. We construct it by 1) first interleaving the input feature maps with zeros, and then 2) convolving them with a customized low-pass filter with adjustable bandwidth (Fig. 2). For filter design, we adopt the Kaiser-Bessel window [18] as it offers explicit control over the tradeoffs between passband ripple and stopband attenuation. The Kaiser window is defined as

$$w(n) = I_0(\beta\sqrt{1-(2n/M)^2})/I_0(\beta), -M/2 \leq n \leq M/2, \tag{3}$$

where M is the desired spatial extent of the window, $\beta \geq 0$ is the shape parameter—the higher it is, the greater the stopband attenuation is (and generally the smoother the image is), and I_0 is the zeroth-order modified Bessel function of the first kind. This plug-and-play upsampler can be *inserted in different layers with different M and β hyperparameters*, offering flexible and precise control.

Lipschitz-Regularized Layers. Compared to the non-trainable upsampling that only attenuates signals, the network layer with nonlinearies is the *only* operation capable of *generating new frequencies* [19]. We regularize their Lipschitz constants to control their sensitivity to input variations, which in turn can affect the spectral bias. Formally, a function $f : \mathcal{X} \to \mathcal{Y}$ is said to be Lipschitz continuous if there is a constant $k > 0$ such that $\|f(\mathrm{x}_1) - f(\mathrm{x}_2)\|_\mathrm{p} \leq k\|\mathrm{x}_1 - \mathrm{x}_2\|_\mathrm{p} \quad \forall \mathrm{x}_1, \mathrm{x}_2 \in \mathcal{X}$, where k is the Lipschitz constant that bounds how fast f can change globally w.r.t. input perturbations. Spectral bias towards low frequencies favors functions with small Lipchitz constants.

Instead of upper bounding the Lipschitz constants of the network layers to pre-defined and manually chosen values as in [39], we make the per-layer Lipschitz bounds learnable and regularize their magnitudes during optimization.

The Lipschitz constant of a convolutional layer is bounded by the operator norm of its weight matrix [10]. To bound a convolutional layer to a specific Lipschitz constant k, the layer with m input channels, c output channels and kernels of size $w \times h$ is first reshaped to a 2-D matrix $W^\ell \in \mathbb{R}^{n \times cwh}$, and normalized as

$$\tilde{W}_\ell = \frac{W_\ell}{\max(1, \frac{\|W_\ell\|_\mathrm{p}}{\texttt{SoftPlus}(k_\ell)})}, \tag{4}$$

where k_ℓ is a learnable Lipschitz constant for each layer, $\|\cdot\|_\mathrm{p}$ is chosen as the ℓ_∞ norm and $\texttt{SoftPlus}(k_\ell) = \ln(1 + \exp(k_\ell))$ ensures the learned Lipschitz bounds are non-negative. Such formulation *only* normalizes W_ℓ if its matrix norm is larger than the learned Lipschitz constraint during training. Integrating the ultimate Lipschitz regularization into Eq. (2), our regularized training objective is

$$\min_{\Theta, K} \mathcal{L}(\mathbf{y}; \mathbf{AG}_{\mathbf{\Theta}}(\mathbf{z})) + \lambda \sum_{l=1}^{L} \texttt{SoftPlus}(\mathbf{k}_\ell)^2 \tag{5}$$

where $\mathbf{K}$ is a collection of per-layer learnable Lipschitz constant k_ℓ jointly optimized with the network parameters, and λ controls the granularity of smoothness.

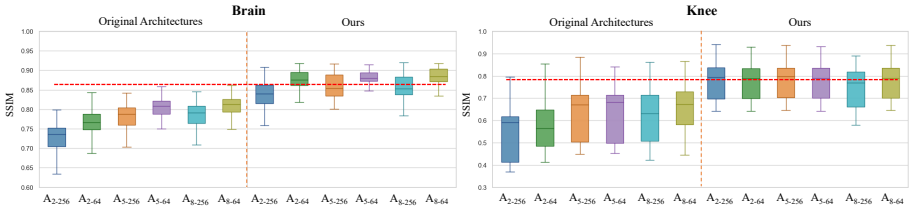


Fig. 5: Our approach significantly minimizes the performance gaps among architectures with various depths $\{2, 5, 8\}$ and widths $\{64, 256\}$.

4 Experiments

4.1 Setup and Datasets

We first (1) validate the effectiveness of the proposed methods in enhancing the performance of untrained networks across various architectural configurations, especially those originally underperforming. We then (2) benchmark the enhanced versions of those *compact* architectures against established supervised and self-supervised methods on both in-domain and out-of-domain datasets in terms of accuracy and efficiency. We also (3) compare our methods with self-validation-based early stopping [46] on overcoming overfitting, and show that it is complementary to our approach by further shortening the reconstruction time. Finally, we demonstrate the utility of our methods in (4) general image inpainting and denoising tasks and (5) perform spectral bias analysis on all evaluated tasks.

The MRI experiments were performed on two publicly available datasets: the multi-coil knee and brain MRI images from fastMRI database [23], and multi-coil knee MRI images from Stanford 3D FSE knee dataset [17]. The fully-sampled k-space data was retrospectively masked by selecting 25 central k-space lines along with a uniform undersampling at outer k-space, achieving the standard $4\times$ acceleration. For training the supervised baseline, the knee training set consists of 367 PD and PDFS slices and the brain training set consists of 651 slices with a mixture of T1 and T2 weighted images. 50 knee slices and 50 brain slices were sampled from the respective multi-coil validation datasets for evaluation.

4.2 Implementation Details

Without loss of generality, the base architectures considered in our work are N-level encoder-decoder architectures with full skip connections. The architectures are isotropic with the same width and kernel size throughout the layers. All evaluated architectures are trained for 3k iterations using mean absolute error and Adam optimizer [20] with a learning rate of 0.008. Unless otherwise specified, the results at the last iteration are reported. The input is drawn from a uniform distribution $z \sim \mathcal{U}(0, 1)$. The filter size of the Gaussian blur was set to 5 and the sigma value was randomly sampled from $[0.5, 2.0]$ for every slice. M and β for the Kaiser-based upsamplers are chosen to be $\{15 \times N - 1, 5\}$ and $\{5 \times N\}$ for

Table 2: Effectiveness of the methods in bridging performance gaps among different architectures, evaluated on fastMRI **brain** datasets. Bandlimited inputs achieved by Fourier features ($L = 4$ or 8) or Gaussian blur along with Lipschitz regulariation improve all architectures, especially the shallower. The proposed Kaiser-based upsampling dramatically improves the deeper architectures. All architectures end up with similarly high performance. The best and the second-best are highlighted.

Regularizers	$\mathbf{A_{2\text{-}256}}$	$\mathbf{A_{2\text{-}64}}$	$\mathbf{A_{5\text{-}256}}$	$\mathbf{A_{5\text{-}64}}$	$\mathbf{A_{8\text{-}256}}$	$\mathbf{A_{8\text{-}64}}$	$\mathbf{A_{2\text{-}256}}$	$\mathbf{A_{2\text{-}64}}$	$\mathbf{A_{5\text{-}256}}$	$\mathbf{A_{5\text{-}64}}$	$\mathbf{A_{8\text{-}256}}$	$\mathbf{A_{8\text{-}64}}$
	PSNR ↑						SSIM ↑					
w/o. Reg. (Plain)	29.08	29.41	31.15	31.42	31.27	31.68	0.729	0.761	0.782	0.801	0.784	0.807
TV	29.22	29.61	31.26	31.37	31.32	31.64	0.735	0.764	0.785	0.802	0.787	0.807
Lipschitz Reg.	30.92	29.73	31.47	32.11	31.50	32.03	0.795	0.766	0.792	0.812	0.800	0.820
Fourier features ($L = 16$)	30.57	30.49	31.57	31.77	31.77	32.09	0.786	0.788	0.794	0.813	0.799	0.819
Fourier features ($L = 8$)	31.42	31.98	31.82	32.42	31.60	32.45	0.804	0.833	0.799	0.831	0.795	0.834
Fourier features ($L = 4$)	31.92	32.59	31.87	32.80	31.71	32.86	0.840	0.863	0.799	0.848	0.793	0.844
Gaussian blurred	33.34	32.67	32.14	32.66	32.03	32.92	0.870	0.866	0.811	0.849	0.825	0.849
Gauss. + Lips.	32.90	33.12	32.08	32.83	31.70	33.14	0.855	0.870	0.815	0.851	0.805	0.849
Gauss. + Lips. + Kaiser Up.	32.50	33.10	33.00	33.21	33.09	33.85	0.836	0.874	0.857	0.876	0.858	0.885

Table 3: Evaluated on fastMRI **knee** datasets.

Regularizers	$\mathbf{A_{2\text{-}256}}$	$\mathbf{A_{2\text{-}64}}$	$\mathbf{A_{5\text{-}256}}$	$\mathbf{A_{5\text{-}64}}$	$\mathbf{A_{8\text{-}256}}$	$\mathbf{A_{8\text{-}64}}$	$\mathbf{A_{2\text{-}256}}$	$\mathbf{A_{2\text{-}64}}$	$\mathbf{A_{5\text{-}256}}$	$\mathbf{A_{5\text{-}64}}$	$\mathbf{A_{8\text{-}256}}$	$\mathbf{A_{8\text{-}64}}$
	PSNR ↑						SSIM ↑					
w/o. Reg. (Plain)	27.18	27.62	29.16	29.23	28.98	29.35	0.541	0.575	0.628	0.640	0.625	0.644
TV	28.25	27.85	29.33	29.57	29.54	30.01	0.588	0.592	0.635	0.651	0.645	0.687
Lipschitz Reg.	28.41	29.21	29.17	29.79	29.43	30.14	0.601	0.600	0.629	0.651	0.636	0.666
Fourier features ($L = 16$)	28.42	28.97	29.58	30.26	29.76	30.38	0.587	0.622	0.653	0.671	0.661	0.681
Fourier features ($L = 8$)	28.61	29.98	29.86	30.72	29.66	30.89	0.604	0.670	0.669	0.693	0.662	0.703
Fourier features ($L = 4$)	32.02	32.07	29.40	31.13	29.55	31.17	0.775	0.781	0.665	0.718	0.668	0.717
Gaussian blurred	30.87	30.89	30.02	31.24	29.31	30.89	0.739	0.768	0.694	0.748	0.698	0.727
Gaussian blurred + Lips.	31.61	31.93	29.40	31.67	29.82	31.58	0.750	0.776	0.702	0.727	0.697	0.732
Gauss. + Lips. + Kaiser Up.	31.92	31.61	31.78	31.60	31.09	31.73	0.777	0.776	0.778	0.776	0.750	0.768

knee data (N denotes the nth-level), and $\{5 \times N\}$ and $\{5 \times N\}$ for brain data, respectively. λ is set to 1 for the Lipschitz regularizer.

4.3 Effectiveness in Reducing Architectural Sensitivity

Fig. 5 gives a quantitative overview of the substantial improvement give by our approach in architectures with diverse configurations on both knee and brain datasets. The different results of the original architectures also confirm the influences of architectural choices on performance. Notably, before applying our methods, the deeper and narrower architectures tend to perform better than their counterparts (more in appendix). This trend aligns with previous works [3, 6, 7, 43] where these architectures tend to be favored in inpainting-like tasks. Here we identify particularly their counterparts (i.e., $\mathbf{A_{x\text{-}256}}$) as "underperforming" architectures. As will be shown in our spectral bias analysis in Sec. 4.5 and appendix, these underperforming architectures learn high frequencies more quickly (though this may be desired for other tasks [29]) and are more susceptible to overfitting, incurring severe artifacts in the output (Figs. 6 and 10). When applied with our methods, as detailed in Tab. 2 and Tab. 3, a large boost in performance is observed in all architectures, especially $\mathbf{A_{2\text{-}256}}$.

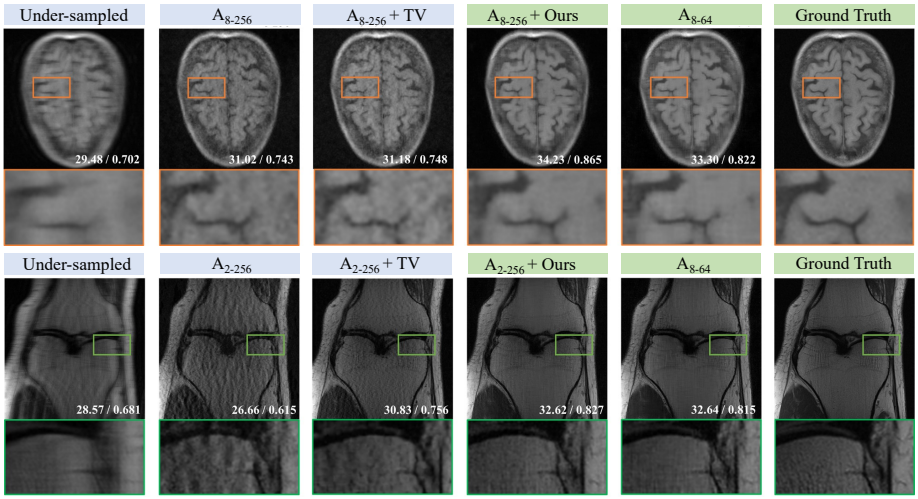


Fig. 6: Our methods enable the underperforming architectures (e.g., $\mathbf{A_{8\text{-}256}}$, $\mathbf{A_{2\text{-}256}}$) to perform similarly to the well-performing architectures (e.g., $\mathbf{A_{8\text{-}64}}$).

We observe that using low-passed inputs via either selected Fourier features or blurring brings the most benefits to the shallower architectures. Better results are achieved when combined with Lipschitz regularization on the layers. On the other hand, deeper architectures benefit more from the Kaiser-based upsampler, which can be seen as performing low-pass filtering on the input feature maps within the network, beyond the initial input layer. We further note that the hyperparameters required for upsampling differ between knee and brain data (Sec. 4.2), with the knee data requiring greater attenuation. This is also consistent with previous findings that they require different numbers of channels for the architectural choices [6]. Our methods greatly alleviate the need for such architectural tuning by instead allowing for the adjustment of a few key hyperparameters.

4.4 Benchmark Results

We adopt two most lightweight architectures as the base models, and compare our regularized network priors with several established MRI reconstruction methods,

Table 4: Quantitative results on fastMRI datasets. Runtime: mean (std) per slice.

Datasets		Supervised UNet	CS-ℓ_1 [16]	ZS-SSL [46]	DIP [43]	Deep Decoder [12]	ConvDecoder [6]	$\mathbf{A_{2\text{-}64}}$ (vanilla)	$\mathbf{A_{8\text{-}64}}$ (vanilla)	$\mathbf{A_{2\text{-}64}}$ (Ours)	$\mathbf{A_{8\text{-}64}}$ (Ours)
Brain	PSNR ↑	33.35	29.91	**34.39**	31.15	26.97	31.81	29.42	31.68	33.10	33.85
	SSIM ↑	**0.889**	0.773	0.878	0.782	0.747	0.800	0.761	0.807	0.874	0.885
Knee	PSNR ↑	31.15	28.23	32.00	29.16	27.21	29.59	27.62	29.35	**32.07**	31.73
	SSIM ↑	0.776	0.633	0.773	0.628	0.687	0.655	0.575	0.644	**0.781**	0.768
GFLOPS ↓		99.24	–	5461.6	615.72	82.82	699.94	38.42	40.94	62.36	68.38
Runtime (mins) ↓		0.002 (0.00003)	–	64.8 (20.18)	14.0 (0.61)	6.6 (0.63)	8.2 (0.35)	5.4 (0.47)	10.5 (0.62)	6.6 (0.58)	12.3 (0.65)

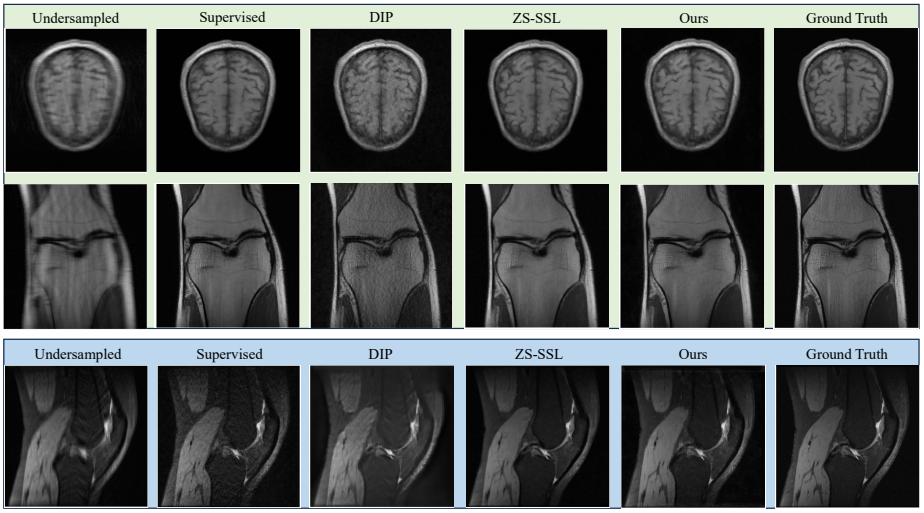


Fig. 7: Qualitative evaluations. In-domain : fastMRI. Out-of-domain : Stanford FSE.

including a supervised baseline, a state-of-the-art self-supervised method based on unrolling networks (`ZS-SSL`) [46] and underparameterized untrained networks (`ConvDecoder` [6], `Deep Decoder` [12]). Visual comparisons are presented in Fig. 7.

Comparisons with state-of-the-arts. Our method enables the previously underperforming architectures to match the performance of `ZS-SSL` and that of the supervised UNet on fastMRI knee data (Tab. 4), and surpass the trained UNet on out-of-domain Stanford 3D FSE data (Tab. 5), demonstrating their advantages in generalizable reconstruction. Our enhanced networks also clearly outperform other lightweight untrained networks, i.e., `ConvDecoder` and `Deep Decoder`, which are designed to prevent overfitting. `ZS-SSL` is an unrolling method where the network, i.e., a ResNet [42], is adopted as a denoiser. Our enhanced network priors achieve comparable performance than `ZS-SSL` while being orders of magnitude faster, thanks to the more efficient base models.

Table 5: Out-of-domain evaluation among supervised and untrained methods and comparisons with a **self-validation-based early stopping** strategy [46].

	Method	**In-domain**		**Out-domain**		Runtime (mean±std)	
		PSNR	SSIM	PSNR	SSIM	Train	Inference
Trained	U-Net	31.16	0.776	29.16	0.724	∼1.5 days	0.1 ± 0.003 sec
Untrained	CS-ℓ_1 [16]	28.23	0.633	22.46	0.407	–	–
	ZS-SSL [46]	32.00	0.773	**31.74**	**0.805**	–	26.1 ± 3.5 mins
	DIP [43]	29.16	0.628	28.89	0.664	–	9.2 ± 0.3 mins
	$\mathbf{A_{2\text{-}64}}$	27.62	0.575	26.03	0.550	–	5.5 ± 0.1 mins
	$\mathbf{A_{2\text{-}64}}$ + Early Stopped	29.59	0.695	27.59	0.641	–	0.2 ± 0.2 mins
	$\mathbf{A_{2\text{-}64}}$ (Ours)	**32.07**	**0.781**	31.43	0.790	–	6.4 ± 0.4 mins
	$\mathbf{A_{2\text{-}64}}$ (Ours) + Early Stopped	31.97	0.776	31.30	0.800	–	0.3 ± 0.1 mins

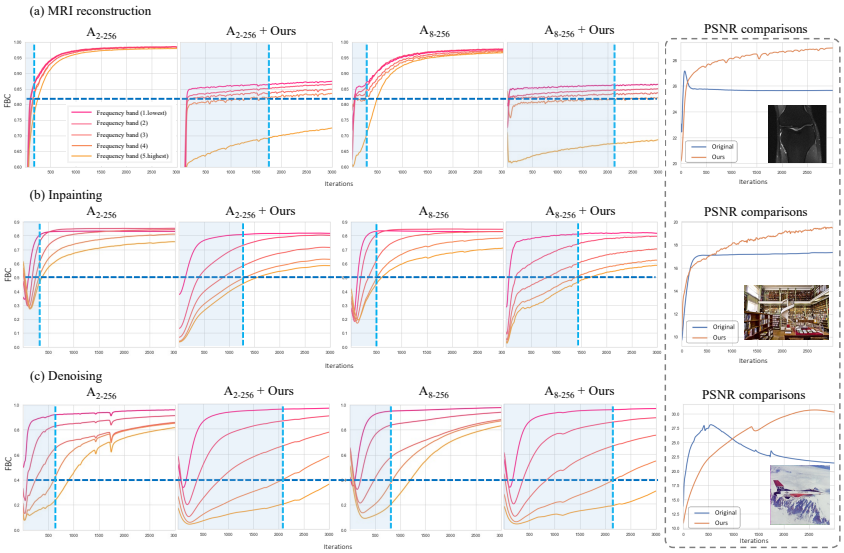


Fig. 9: Measurement of spectral bias. Underperforming architectures (e.g., $A_{2\text{-}256}$, $A_{8\text{-}256}$) tend to learn high frequencies hastily, overfit more easily, and extrapolate poorly. Our method effectively mitigates these shortcomings.

Comparisons with early stopping (ES). ZS-SSL uses a self-validation strategy for early stopping whereas our method does not necessitate ES. Our method not only alleviates overfitting but more importantly, enhances the inter/extrapolation capabilities of the underperforming architectures, leading to higher peak performance than the original scheme (Tab. 5). This cannot be achieved even with the best ES strategy, as ES only halts the training near the peak PSNR for a given architecture, but it cannot fundamentally improve the underperforming architectures whose peak PSNR remains subpar (Fig. 8). We show in Tab. 5 that self-validation based ES can be integrated into our approach to further shorten the reconstruction time from 6 mins→0.3 mins. More examples are in Suppl. Fig. 8.

Fig. 8: 8× undersampling

4.5 Spectral Bias Analysis

To examine how the proposed methods influence the frequency bias of the network, we measure the spectral bias using the metric— frequency-band correspondence (FBC) [39], which first calculates the element-wise $|\mathcal{F}(x)|/|\mathcal{F}(y)|$ between the output x and target image y, categorizes it into five frequency bands radially and then computes the per-band averages. Higher values indicate higher cor-

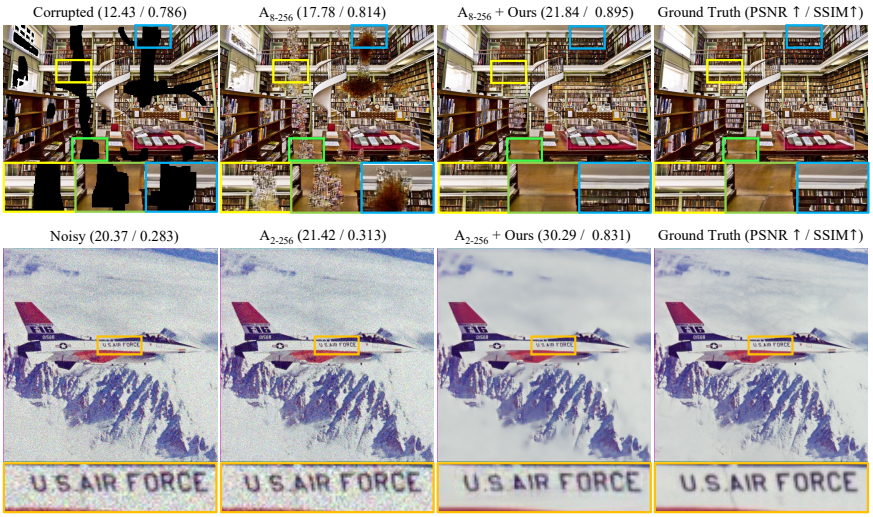


Fig. 10: Experiments on natural image inpainting and denoising ($\sigma = 25$). Our method improves the extrapolation and denoising capabilities of underperforming architectures.

respondence. We trace the evolutions of FBC for $\mathbf{A_{8\text{-}256}}$ and $\mathbf{A_{2\text{-}256}}$ and the corresponding PSNR curves throughout the training iterations in three tasks.

Fig. 9 shows that the original underperforming architectures tend to fit all frequencies more readily, including high frequencies. This is more evident in $\mathbf{A_{2\text{-}256}}$, corresponding to its worst performance among all compared architectures. Our methods substantially delay the convergence of higher frequencies for all three tasks as designed, which decouples the learning of different frequencies. This leads to prolonged denoising effects and enhanced performance for inpainting tasks, including MRI reconstruction, as qualitatively shown in Fig. 10. We speculate that a stronger bias towards lower frequencies helps the model leverage better the spatial information which improves its inter/extrapolation capability. Given that MRI reconstruction resembles an inpainting task for k-space measurements, we expect similar improvements in natural image inpainting shown in Fig. 10 with our regularized network priors for k-space interpolation in MRI experiments.

5 Conclusion

We introduce efficient, architecture-agnostic methods for frequency control over the network priors, offering a novel solution to simultaneously address the key challenges present in untrained image reconstruction. Our approach requires only minimal modifications to the original DIP scheme while achieving significant gains in accruacy and efficiency as evidenced in MRI reconstruction and natural image restoration tasks, making it a stronger zero-shot reconstructior with the potential for seamless integration with other advancements in self-supervised reconstruction.

Acknowledgements

This work was supported in part by the United States National Institutes of Health (NIH) through grants R01CA266702 and R01EB035160.

References

1. Arican, M.E., Kara, O., Bredell, G., Konukoglu, E.: Isnas-dip: Image-specific neural architecture search for deep image prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1960–1968 (2022)
2. Barbano, R., Antorán, J., Leuschner, J., Hernández-Lobato, J.M., Kereta, Ž., Jin, B.: Fast and painless image reconstruction in deep image prior subspaces. arXiv preprint arXiv:2302.10279 (2023)
3. Barbano, R., Leuschner, J., Schmidt, M., Denker, A., Hauptmann, A., Maass, P., Jin, B.: An educated warm start for deep image prior-based micro ct reconstruction. IEEE Transactions on Computational Imaging (2022)
4. Chakrabarty, P., Maji, S.: The spectral bias of the deep image prior. arXiv preprint arXiv:1912.08905 (2019)
5. Chen, Y.C., Gao, C., Robb, E., Huang, J.B.: Nas-dip: Learning deep image prior with neural architecture search. In: European Conference on Computer Vision. pp. 442–459. Springer (2020)
6. Darestani, M.Z., Heckel, R.: Accelerated mri with un-trained neural networks. IEEE Transactions on Computational Imaging **7**, 724–733 (2021)
7. Darestani, M.Z., Liu, J., Heckel, R.: Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing. In: International Conference on Machine Learning. pp. 4754–4776. PMLR (2022)
8. E Woods, R., C Gonzalez, R.: Digital image processing (2008)
9. Fridovich-Keil, S., Gontijo Lopes, R., Roelofs, R.: Spectral bias in practice: The role of function frequency in generalization. Advances in Neural Information Processing Systems **35**, 7368–7382 (2022)
10. Gouk, H., Frank, E., Pfahringer, B., Cree, M.J.: Regularisation of neural networks by enforcing lipschitz continuity. Machine Learning **110**, 393–416 (2021)
11. Hansen, M.S., Kellman, P.: Image reconstruction: an overview for clinicians. Journal of Magnetic Resonance Imaging **41**(3), 573–585 (2015)
12. Heckel, R., Hand, P.: Deep decoder: Concise image representations from untrained non-convolutional networks. arXiv preprint arXiv:1810.03982 (2018)
13. Heckel, R., Soltanolkotabi, M.: Denoising and regularization via exploiting the structural bias of convolutional generators. International Conference on Representation Learning (2020)
14. Ho, K., Gilbert, A., Jin, H., Collomosse, J.: Neural architecture search for deep image prior. Computers & Graphics **98**, 188–196 (2021)
15. Hoffman, J., Roberts, D.A., Yaida, S.: Robust learning with jacobian regularization. arXiv preprint arXiv:1908.02729 (2019)
16. Jaspan, O.N., Fleysher, R., Lipton, M.L.: Compressed sensing mri: a review of the clinical literature. The British journal of radiology **88**(1056), 20150487 (2015)
17. K, E., Am, S., M, L., M, A., M., U.: Creation of fully sampled mr data repository for compressed sensing of the knee

18. Kaiser, J.: Nonrecursive digital filter design using the i0-sinh window function, paper presented at symposium on circuits and systems, inst. of electr. and electron. In Proc. 1974 IEEE International Symposium on Circuits & Systems, pages 20-23, 1974 (1974)
19. Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T.: Alias-free generative adversarial networks. Advances in Neural Information Processing Systems **34**, 852–863 (2021)
20. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Knoll, F., Hammernik, K., Kobler, E., Pock, T., Recht, M.P., Sodickson, D.K.: Assessment of the generalization of learned image reconstruction and the potential for transfer learning. Magnetic resonance in medicine **81**(1), 116–128 (2019)
22. Knoll, F., Murrell, T., Sriram, A., Yakubova, N., Zbontar, J., Rabbat, M., Defazio, A., Muckley, M.J., Sodickson, D.K., Zitnick, C.L., et al.: Advancing machine learning for mr image reconstruction with an open competition: Overview of the 2019 fastmri challenge. Magnetic resonance in medicine **84**(6), 3054–3070 (2020)
23. Knoll, F., Zbontar, J., Sriram, A., Muckley, M.J., Bruno, M., Defazio, A., Parente, M., Geras, K.J., Katsnelson, J., Chandarana, H., et al.: fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. Radiology: Artificial Intelligence **2**(1), e190007 (2020)
24. Korkmaz, Y., Dar, S.U., Yurt, M., Özbey, M., Cukur, T.: Unsupervised mri reconstruction via zero-shot learned adversarial transformers. IEEE Transactions on Medical Imaging (2022)
25. Lingala, S.G., Hu, Y., DiBella, E., Jacob, M.: Accelerated dynamic mri exploiting sparsity and low-rank structure: kt slr. IEEE transactions on medical imaging **30**(5), 1042–1054 (2011)
26. Liu, D., Wang, J., Shan, Q., Smyl, D., Deng, J., Du, J.: Deepeit: deep image prior enabled electrical impedance tomography. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
27. Liu, J., Sun, Y., Xu, X., Kamilov, U.S.: Image restoration using total variation regularized deep image prior. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 7715–7719. IEEE (2019)
28. Liu, Y., Chen, Y., Yap, P.T.: Real-time mapping of tissue properties for magnetic resonance fingerprinting. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 161–170. Springer (2021)
29. Liu, Y., Li, J., Pang, Y., Nie, D., Yap, P.T.: The devil is in the upsampling: Architectural decisions made simpler for denoising with deep image prior. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12408–12417 (2023)
30. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **58**(6), 1182–1195 (2007)
31. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
32. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957 (2018)

33. Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., Sutskever, I.: Deep double descent: Where bigger models and more data hurt. Journal of Statistical Mechanics: Theory and Experiment **2021**(12), 124003 (2021)
34. Nittscher, M., Lameter, M., Barbano, R., Leuschner, J., Jin, B., Maass, P.: Svd-dip: Overcoming the overfitting problem in dip-based ct reconstruction. arXiv preprint arXiv:2303.15748 (2023)
35. Novak, R., Bahri, Y., Abolafia, D.A., Pennington, J., Sohl-Dickstein, J.: Sensitivity and generalization in neural networks: an empirical study. arXiv preprint arXiv:1802.08760 (2018)
36. Qayyum, A., Ilahi, I., Shamshad, F., Boussaid, F., Bennamoun, M., Qadir, J.: Untrained neural network priors for inverse imaging problems: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
37. Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A.: On the spectral bias of neural networks. In: International Conference on Machine Learning. pp. 5301–5310. PMLR (2019)
38. Rosca, M., Weber, T., Gretton, A., Mohamed, S.: A case for new neural network smoothness constraints. In: Zosa Forde, J., Ruiz, F., Pradier, M.F., Schein, A. (eds.) Proceedings on "I Can't Believe It's Not Better!" at NeurIPS Workshops. Proceedings of Machine Learning Research, vol. 137, pp. 21–32. PMLR (12 Dec 2020), `https://proceedings.mlr.press/v137/rosca20a.html`
39. Shi, Z., Mettes, P., Maji, S., Snoek, C.G.: On measuring and controlling the spectral bias of the deep image prior. International Journal of Computer Vision **130**(4), 885–908 (2022)
40. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
41. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems **33**, 7537–7547 (2020)
42. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 114–125 (2017)
43. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9446–9454 (2018)
44. Wang, H., Li, T., Zhuang, Z., Chen, T., Liang, H., Sun, J.: Early stopping for deep image prior. arXiv preprint arXiv:2112.06074 (2021)
45. Xu, Z.Q.J., Zhang, Y., Xiao, Y.: Training behavior of deep neural network in frequency domain. In: Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part I 26. pp. 264–274. Springer (2019)
46. Yaman, B., Hosseini, S.A.H., Akçakaya, M.: Zero-shot self-supervised learning for mri reconstruction. International Conference on Learning Representations (2022)
47. Yang, J., Pavone, M., Wang, Y.: Freenerf: Improving few-shot neural rendering with free frequency regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8254–8263 (2023)
48. Yu, T., Hilbert, T., Piredda, G.F., Joseph, A., Bonanno, G., Zenkhri, S., Omoumi, P., Cuadra, M.B., Canales-Rodríguez, E.J., Kober, T., et al.: Validation and generalizability of self-supervised image reconstruction methods for undersampled mri. arXiv preprint arXiv:2201.12535 (2022)