Image Reconstruction Via Autoencoding Deep Image Prior

Ismail Alkhouri*, Shijun Liang*,
Evan Bell, Qing Qu, Rongrong Wang, and Saiprasad Ravishankar

December 2024

NEURAL INFORMATION PROCESSING SYSTEMS

College of Engineering
MICHIGAN STATE UNIVERSITY

MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Overview

- **Deep Image Prior** (DIP) [1] has emerged as an effective unsupervised image reconstructor, requiring only untrained networks.

- **Main Challenge:** DIP vulnerability to noise overfitting artifacts.

- **Approaches to mitigate Noise Overfitting**:
    1. Regularization [2].
    2. Early Stopping [3].
    3. Network Pruning [4].

[1] Deep Image Prior. CVPR, 2018.
[2] Robust self-guided deep image prior. ICASSP, 2023.
[3] Early stopping for deep image prior. TMLR, 2023
[4] Optimal Eye Surgeon: Finding Image Priors through Sparse Generators at Initialization. ICML, 2023.

Contributions

Building upon

1. An insight on the DIP network input,
2. The gradual denoising process of diffusion models,

we introduce Autoencoding Sequential Deep Image Prior (**aSeqDIP**), which updates based on objective functions that consist of:

1. Input-adaptive data consistency term, and
2. Autoencoding term for noise overfitting mitigation.

- **Evaluation**: aSeqDIP vs. DIP and DM baselines in terms of
  1. reconstruction quality,
  2. run-time, and
  3. robustness to noise overfitting artifacts.

Preliminaries

- **Task:** Recover an image $\mathbf{x}^* \in \mathbb{R}^n$ from measurements $\mathbf{y} \in \mathbb{R}^m$, where $m \leq n$, governed by forward model $\mathbf{A}$.
- **DIP:** uses a random fixed network input $\mathbf{z}$ and reconstructs image $\hat{\mathbf{x}}$ by

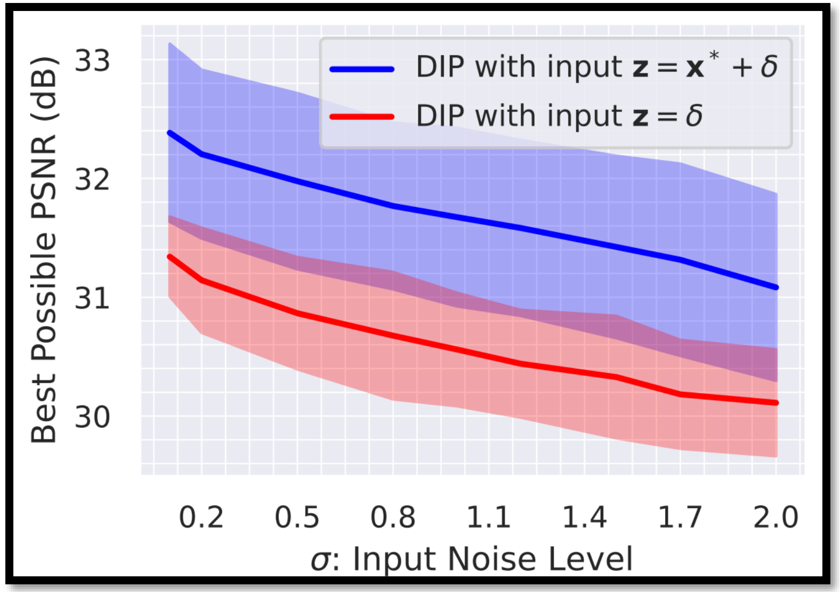
$$\hat{\theta} = \arg\min_{\theta} || \mathbf{A} f_\theta(\mathbf{z}) - \mathbf{y}||_2^2, \quad \hat{\mathbf{x}} = f_{\hat{\theta}}(\mathbf{z})$$

- **Problem**: Selecting the number of iterations to optimize the above objective.
- **Reason:** The network eventually fits the noise present in $\mathbf{y}$ or could fit to undesired images based on the null space of $\mathbf{A}$.

Autoencoding Sequential DIP (aSeqDIP)

- **Motivation:** How does employing a noisy version of the ground truth image, as the fixed input to the Vanilla DIP objective, affect performance?
- Consider the MRI reconstruction task with $\mathbf{y} \approx \mathbf{A}\mathbf{x}^*$. Let the input to f in DIP be

$$\mathbf{z} = \mathbf{x}^* + \boldsymbol{\delta}, \qquad \boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$



Average over 8 MRI scans where the input $\mathbf{z}$ is either a perturbed version of the ground truth or pure noise, controlled by σ (x-axis).

Observation: The proximity of the DIP network input to the ground truth correlates with the quality of the reconstruction.

Question: *Can we develop an <u>input-adaptive</u> DIP method that mitigates noise overfitting?*

Autoencoding Sequential DIP (aSeqDIP)

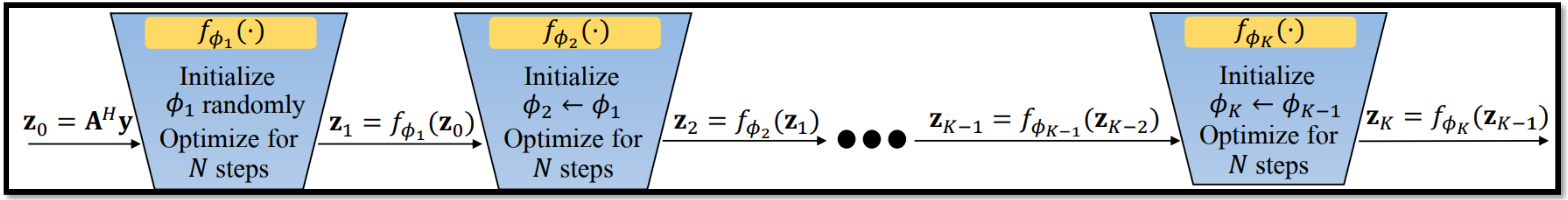
- **Method:** Consider a U-Net $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ whose weights are updated sequentially: ϕ_k with $k \in [K]$ and $[K] = \{1, \ldots, K\}$. Each f_{ϕ_k} takes input $\mathbf{z}_k$ and outputs $f_{\phi_k}(\mathbf{z}_k)$.

Parameters update

$$\phi_k \leftarrow \arg\min_{\phi_k} \underbrace{||\mathbf{A} f_{\phi_k}(\mathbf{z}_{k-1}) - \mathbf{y}||_2^2}_{\text{Data Consistency}} + \underbrace{\lambda || f_{\phi_k}(\mathbf{z}_{k-1}) - \mathbf{z}_{k-1}||_2^2}_{\text{Autoencoder Regularization}}$$

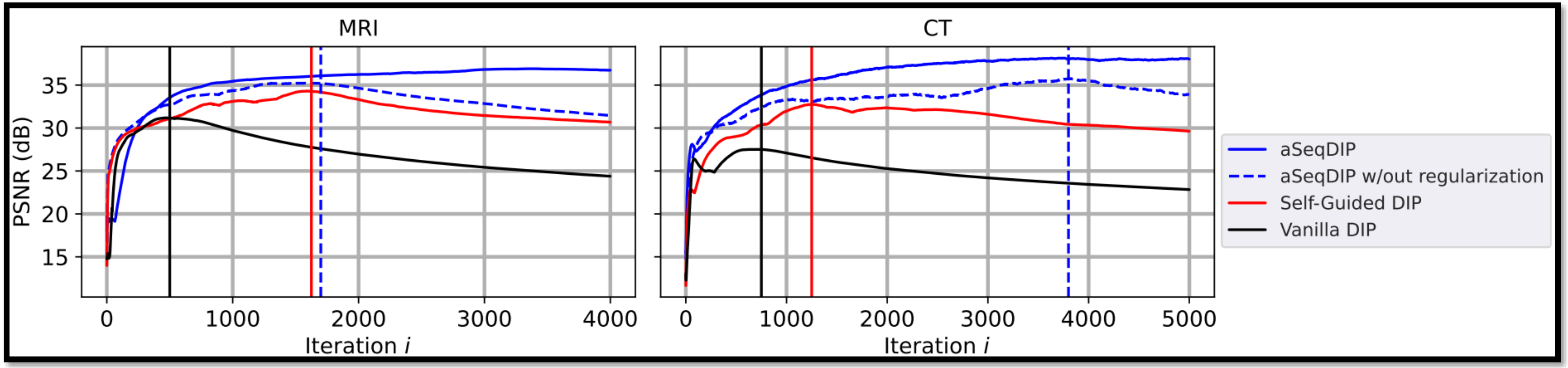
Network Input update

$$\mathbf{z}_k \leftarrow f_{\phi_k}(\mathbf{z}_{k-1})$$



Experimental Results

- **Robustness To Noise Overfitting**:



Average PSNR results w.r.t. iteration $i \in [NK]$ of 20 MRI scans (4x) and 20 CT scans (18 views) to show the impact of the autoencoding term. Vertical lines approximately indicate the start of the PSNR decay for every case

Experimental Results

- **Main Results**

Task	Setting
MRI	Ax $\in \{4x, 8x\}$
CT	views $\in \{18, 30\}$
Denoising	$\sigma_d \in \{15, 30\}$
In-painting	HIAR $\in \{0.1, 0.25\}$
Deblurring	BKSE [3]

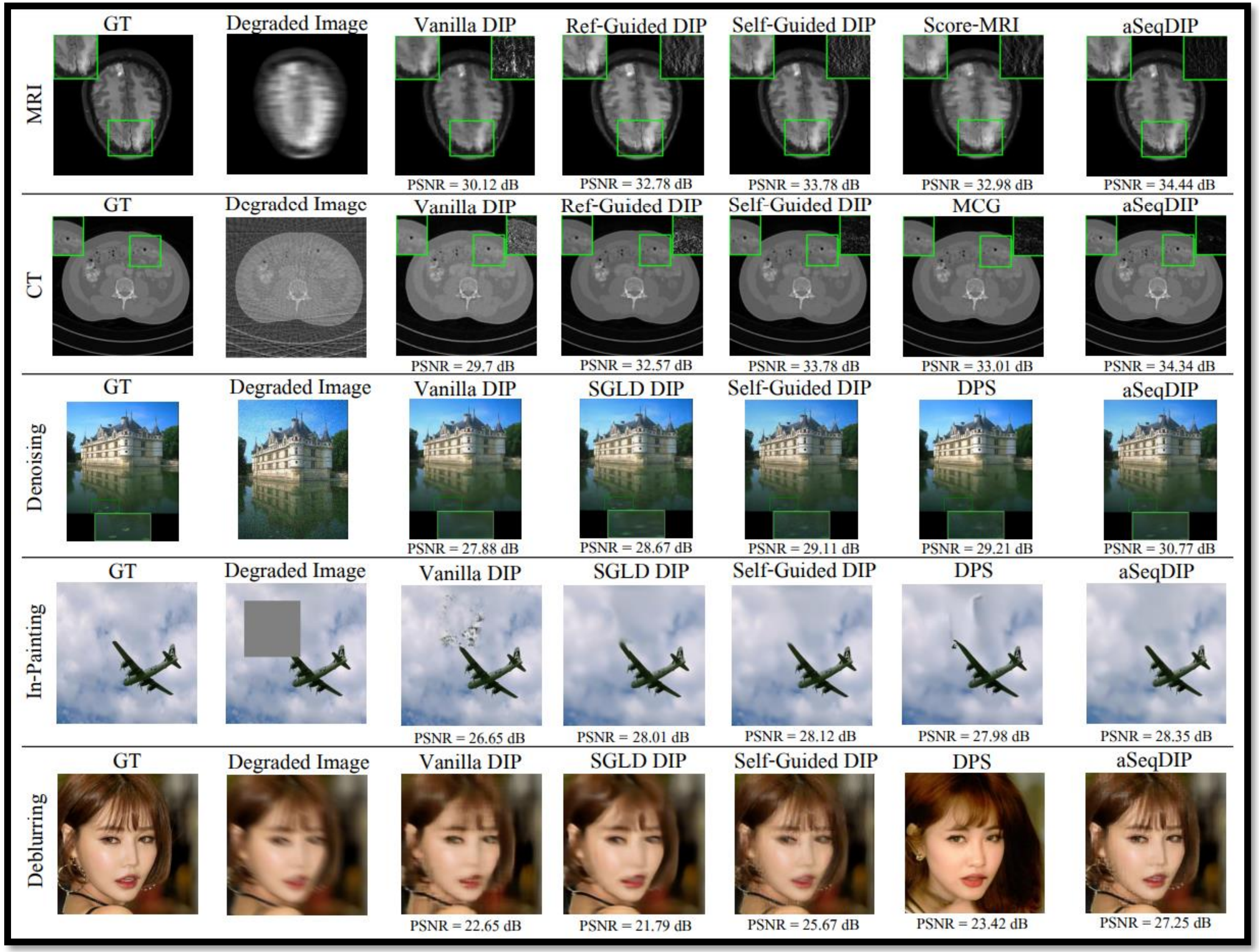
Task	Method	Data Independency	PSNR (dB) (↑) (Setting 1, Setting 2)	SSIM $\in [0,1]$ (↑) (Setting 1, Setting 2)	Run-time (↓) (minutes)
	Score-MRI	×	(31.51±0.45, 29.61±0.44)	(0.891±0.012, 0.862±0.014)	6.2±0.12
	Ref-Guided DIP	×	(33.17±0.27, 30.23±0.24)	(0.912±0.021, 0.873±0.016)	2.5±0.2
	TV-DIP	✓	(30.52±0.25, 29.20±0.37)	(0.872±0.022, 0.852±0.022)	2.5±0.1
MRI	ES-DIP	✓	(31.02±0.34, 29.44±0.45)	(0.882±0.031, 0.858±0.028)	1.56±0.34
	Vanilla DIP	✓	(30.21±0.42, 28.75±0.33)	(0.865±0.02, 0.842±0.022)	**1.5±0.12**
	Self-Guided DIP	✓	(33.6±0.23, 30.75±0.25)	(0.922±0.008, 0.874±0.006)	4.5±0.67
	aSeqDIP (Ours)	✓	(**34.08±0.41**, **31.34±0.47**)	(**0.929±0.008**, **0.887±0.009**)	2.2±0.12
	MCG	×	(32.82±0.52, 31.35±0.49)	(0.912±0.08, 0.852±0.09)	6.4±0.2
	FBP	✓	(22.92±0.22, 19.52±0.32)	(0.75±0.021, 0.68±0.023)	**0.2±0.01**
CT	Ref-Guided DIP	×	(31.21±0.24, 28.31±0.42)	(0.892±0.023, 0.842±0.021)	2.5±0.42
	Vanilla DIP	✓	(26.21±0.12, 24.31±0.34)	(0.791±0.021, 0.772±0.012)	1.5±0.21
	Self-Guided DIP	✓	(33.95±0.32, 31.95±0.32)	(0.918±0.02, 0.872±0.031)	4.5±0.56
	aSeqDIP (Ours)	✓	(**34.88±0.36**, **33.09±0.39**)	(**0.941±0.026**, **0.92±0.022**)	2.2±0.42
	DPS	×	(31.02±0.25, 28.2±0.31)	(0.912±0.02, 0.882±0.021)	2.5±0.17
	Vanilla DIP	✓	(30.48±0.28, 27.84±0.32)	(0.905±0.021, 0.871±0.030)	**1.5±0.22**
	SGLD DIP	✓	(30.58±0.34, 28.12±0.42)	(0.908±0.021, 0.877±0.017)	3.2±0.24
Denoising	TV-DIP	✓	(30.57±0.31, 28.47±0.26)	(0.914±0.022, 0.882±0.014)	2.5±0.24
	Rethinking-DIP	✓	(30.98±0.31, 28.67±0.25)	(0.912±0.02, 0.887±0.03)	2.5±0.34
	ES-DIP	✓	(31.11±0.23, 28.12±0.41)	(0.914±0.017, 0.886±0.024)	1.45±0.44
	Self-Guided DIP	✓	(31,21±0.26, 28.31±0.35)	(0.916±0.02, 0.891±0.03)	3.5±0.45
	aSeqDIP (Ours)	✓	(**31.51±0.34**, **28.97±0.44**)	(**0.926±0.021**, **0.908±0.031**)	2.4±0.45
	DPS	×	(23.9±0.45, 22.03±0.36)	(0.817±0.023, 0.762±0.021)	2.5±0.3
	Vanilla DIP	✓	(22.56±0.31, 21.32±0.67)	(0.754±0.023, 0.721±0.012)	**1.5±0.35**
In-Painting	SGLD DIP	✓	(23.09±0.55, 21.41±0.45)	(0.772±0.023, 0.732±0.041)	2.5±0.45
	TV-DIP	✓	(22.87±0.45, 21.64±0.51)	(0.774±0.04, 0.742±0.042)	2.5±0.31
	ES-DIP	✓	(23.33±0.44, 21.89±0.28)	(0.781±0.034, 0.745±0.041)	1.25±0.55
	Self-Guided DIP	✓	(23.84±0.43, 21.78±0.52)	(0.792±0.042, 0.752±0.064)	3.5±0.45
	aSeqDIP (Ours)	✓	(**24.56±0.45**, **22.57±0.47**)	(**0.838±0.051**, **0.778±0.045**)	2.4±0.54
	DPS	×	(23.40±0.56)	(0.776±0.032)	2.24±0.65
Deblurring	SGLD DIP	✓	(19.80±0.43)	(0.720±0.03)	3.24±0.55
	Self-Guided DIP	✓	(20.34±0.55)	(0.732±0.025)	3.4±1.02
	aSeqDIP (Ours)	✓	(**23.89±0.40**)	(**0.792±0.033**)	2.5±0.78

Experimental Results

- **Visualizations:**

Main Message: aSeqDIP

- is robust to overfitting artifacts.
- is applicable to many (non)linear tasks.
- outperforms all DIP-based methods in terms of reconstruction Quality.
- is on-par with leading DM-based methods.



Thank you!

Code

Paper

Acknowledgments: This work was supported in part by the National Science Foundation (NSF) grants CCF-2212065, CCF-2212066, and BCS-2215155.