Bagged Deep Image Prior for Recovering Images in the Presence of Speckle Noise

Xi Chen[1] Zhewen Hou[2] Christopher A. Metzler[3]
Arian Maleki[2] Shirin Jalali[1]

[1] Rutgers University, [2] Columbia University, [3] UMD

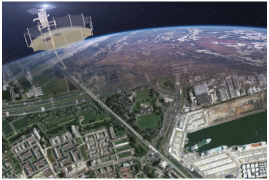
ICML, Vienna, Austria, July 2024

Coherent imaging

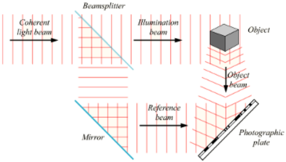
Various imaging systems employ coherent light:

- Optical coherence tomography (OCT)
- Synthetic aperture radar (SAR)
- Inverse synthetic aperture radar (ISAR)
- Digital holography

One of the key challenges is multiplicative noise.



SAR imaging



Digital holography

Coherent imaging: mathematical model

$$y = AXw + z$$

- $X = \mathrm{diag}(x) \in \mathbb{R}^{n \times n}$, where $x \in \mathbb{R}^n$ is the desired signal.
- $w \in \mathbb{R}^n$: signal-dependent speckle noise.
- $A \in \mathbb{R}^{m \times n}$: known sensing matrix, $m < n$.
- $z \in \mathbb{R}^n$: additive white Gaussian noise.

Goal: Recover x from measurements $y = AXw + z$.

To improve performance: acquire multiple independent looks, i.e., recording x under various realizations of the noise process:

$$y_\ell = AXw_\ell + z_\ell$$

MLE of multiple compressive looks

We derive the maximum likelihood estimator (MLE) of multi-look compressive measurements:

$$\hat{x} = \underset{X=\mathrm{diag}(x):x\in\mathscr{C}}{\arg\min} \, f_L(x)$$

$$f_L(x) = \log\det(AX^2A^T) + \frac{1}{L}\sum_{\ell=1}^{L} y_\ell^T (AX^2A^T)^{-1} y_\ell$$

Projected Gradient Descent for solving MLE-based recovery:

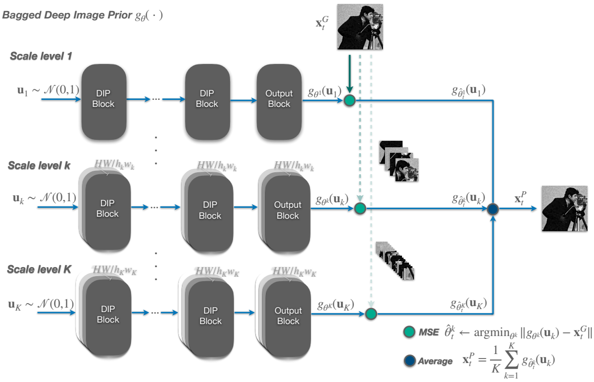
- Initialize X_0.
- For $t = 1, 2, \ldots$:
  i. Gradient descent: $S_{t+1} = X_t - \mu \nabla f_L(X_t)$
  ii. Projection: $X_{t+1} = \arg\min_\theta \| g_\theta(u) - S_{t+1} \|_2^2$

Bagged Deep Image Prior (Bagged-DIP)

Deep image prior (DIP) hypothesis

Choose $u \sim \mathcal{N}(0, I_d)$. $x \in \mathcal{Q}$ can be effectively expressed as $x \approx g_\theta(u)$

Bagging idea: Average over several low-bias and hopefully weakly dependent estimates (DIP outputs) yields a lower-variance estimate.



Performance of MLE as a function of m, n, L

Sensor measurements:

$$y_\ell = AXw_\ell, \quad \ell = 1, 2, \ldots, L$$

Recovery method:

$$\hat{x} = \underset{x: x = g_\theta(u)}{\operatorname{argmin}} f_L(x)$$

Theorem

Suppose i.i.d. $A_{ij} \sim \mathcal{N}(0,1)$, $m < n$ and $g_\theta(u)$, as function of $\theta \in [-1,1]^k$, is 1-Lipschitz, we have

$$\frac{1}{n}\|\hat{x} - x\|_2^2 = O\left(\frac{\sqrt{k\log n}}{m} + \frac{n\sqrt{k\log n}}{m\sqrt{Lm}}\right).$$

with probability $1 - O(e^{-\frac{m}{2}} + e^{-\frac{Ln}{8}} + e^{-k\log n} + \mathrm{e}^{k\log n - \frac{n}{2}})$

Newton-Schulz matrix inversion approximation

In PGD, gradient computation in each iteration t involves $m \times m$ matrix $B_t = AX_t^2 A^T$ inversion:

$$\frac{\partial f_L}{\partial x_j} = 2x_j\Big(a_j^T {B_t}^{-1} a_j - \frac{1}{\sigma_w^2 L}\sum_{\ell=1}^{L}\left(a_j^T {B_t}^{-1} y_\ell\right)^2\Big)$$

Instead of computing the matrix inverse directly, the iterations of Newton-Schulz for finding $(B_t)^{-1}$ is given by

$$M^k = M^{k-1} + M^{k-1}(I - B_t M^{k-1})$$

where M^k is the approximation of $(B_t)^{-1}$ at iteration k. It starts with $M^0 = B_{t-1}^{-1}$, which is the matrix inverse from previous GD iteration.

Empirical results align with theoretical bounds

m/n	#looks	Barbara	Peppers	House	⋯	Average
12.5%	25	19.91/0.443	19.70/0.385	20.15/0.377	⋯	19.24/0.406
	50	20.90/0.567	21.69/0.535	22.27/0.531	⋯	20.83/0.538
	100	21.84/0.633	22.41/0.657	23.96/0.624	⋯	21.78/0.612
25%	25	23.57/0.586	23.17/0.547	24.25/0.520	⋯	22.86/0.549
	50	25.38/0.689	25.12/0.691	26.84/0.652	⋯	24.95/0.672
	100	26.26/0.748	26.14/0.759	28.33/0.717	⋯	26.24/0.745
50%	25	27.30/0.759	27.02/0.724	28.56/0.697	⋯	27.21/0.740
	50	28.67/0.816	28.52/0.804	30.30/0.762	⋯	28.78/0.818
	100	29.40/0.843	29.21/0.849	31.61/0.815	⋯	29.78/0.856

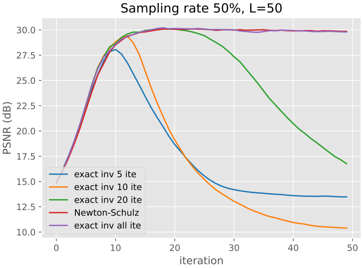
Sharpness of the bound, where the dominant term is $\frac{n\sqrt{k\log n}}{m\sqrt{Lm}}$.

- The decay in terms of m is $m^{3/2}$, and in terms of L is $L^{1/2}$.
  - PSNR gain with double m: theoretical value is $15\log 2 \approx 4.5$dB, empirical value is 3.99dB.
  - PSNR gain with double L: theoretical value is $5\log 2 \approx 1.5$dB, empirical value is 1.42dB.

Efficiency and effectiveness of Newton-Schulz

Evaluation of Newton-Schulz:

- Newton-Schulz is effective: virtually identical to PGD with the exact inverse.
- Newton-Schulz is necessary: exact inverse only for certain # iterations diverges.



Time (in sec) required for exact matrix inversion and its Newton-Schulz approximation in PGD step.

Image size	**32 × 32**	**64 × 64**	**128 × 128**
GD w/ Newton-Schulz	~ 7e-5	~ 8e-5	~ 1e-4
GD w/o Newton-Schulz	~ 0.3	~ 1.2	~ 52.8

Improvements provided by bagging DIPs

Evaluation of the bagging idea:

- Bagged-DIP is effective:

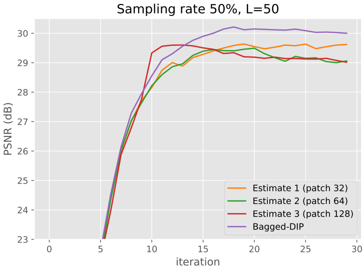
  Bagged-DIP has offered $0.5-1$dB over the three estimates it has combined.

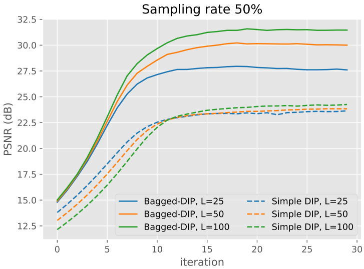
  - The gain is expected to increase when the number of estimates K increases.

- Bagged-DIP is more robust:

  Bagged-DIP overcomes the bottleneck caused by the simple structured DIP when L increases.

  - The theoretical gain is ≈ 1.5dB when L doubles.





Main contributions

Theoretical contribution:

- First existing MLE-based recovery error bound $\|\hat{x} - x\|_2^2$ in terms of parameters (m, n, L, k).

Algorithmic contribution:

- Bagged Deep Image Prior projection: $x = g_\theta(u)$
  - Bagging of independent DIPs to provide more robust and effective projection.
- Newton Schulz matrix inversion approximation
  - Efficient matrix inversion approximation in PGD to avoid exact large matrix inversion.

Thanks you!