The Spectral Bias of the Deep Image Prior

UMass Amherst
College of Information and Computer Sciences

Prithvijit Chakrabarty
pchakrabarty@cs.umass.edu

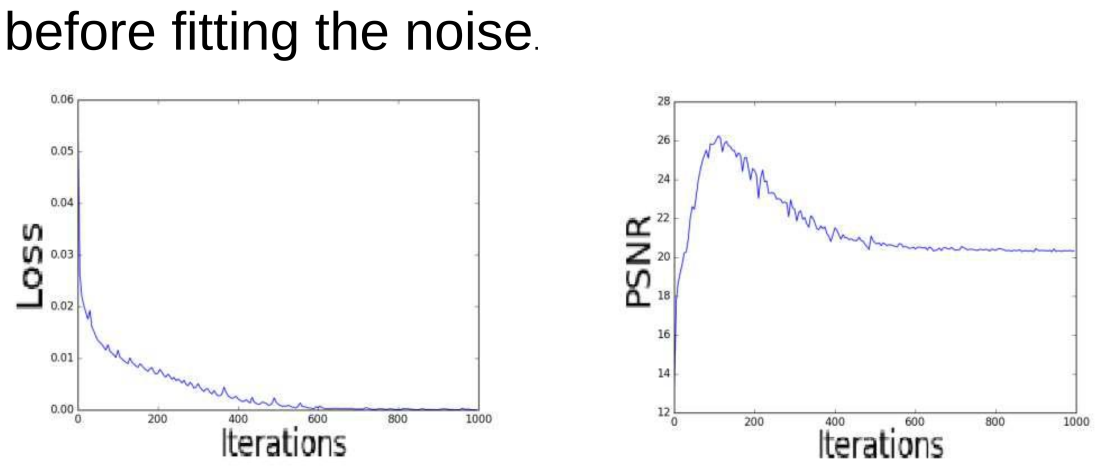
Subhransu Maji
smaji@cs.umass.edu

Overview

- We study the Deep Image Prior (DIP) [1] and show:
  - DIP has a spectral bias: the model fits low frequency components of a signal first.
  - DIP can remove additive Gaussian noise, but not low frequency noise on natural images. This is caused by the spectral bias.

The Deep Image Prior

- Given a signal x_0, white noise z, a convolutional encoder-decoder f_θ, DIP studies the following optimization: $\min_\theta ||f_\theta(z) - x_0||^2$
- N steps of gradient descent traces out a trajectory in the parameter space: $\theta^1, \dots \theta^N$. This has a corresponding trajectory T in the output space: $f_{\theta^1}(z), \dots f_{\theta^N}(z)$.
- If N is large, the model with eventually fit x_0 exactly, i.e., $f_{\theta^N}(z) = x_0$.
- If x_0 is a noisy image, T will contain the clean image [1]: the model predicts the clean image before fitting the noise.



DIP as a Gaussian Process

- Convolution layers and non-linearities modify the spatial covariance of the input.
  The input z is white noise. This is a GP with zero mean and is stationary (covariance $K_x(t_1, t_2) = \sigma^2 \mathbf{1}[t_1 == t_2]$).
  The output of the DIP model is also zero mean, and stationary.
- Conditions: (a) the number of input channels and filters go to infinity (b) filters have random normal weights
- This shows a bias towards smooth outputs when the model is initialized. The bias increases with depth.

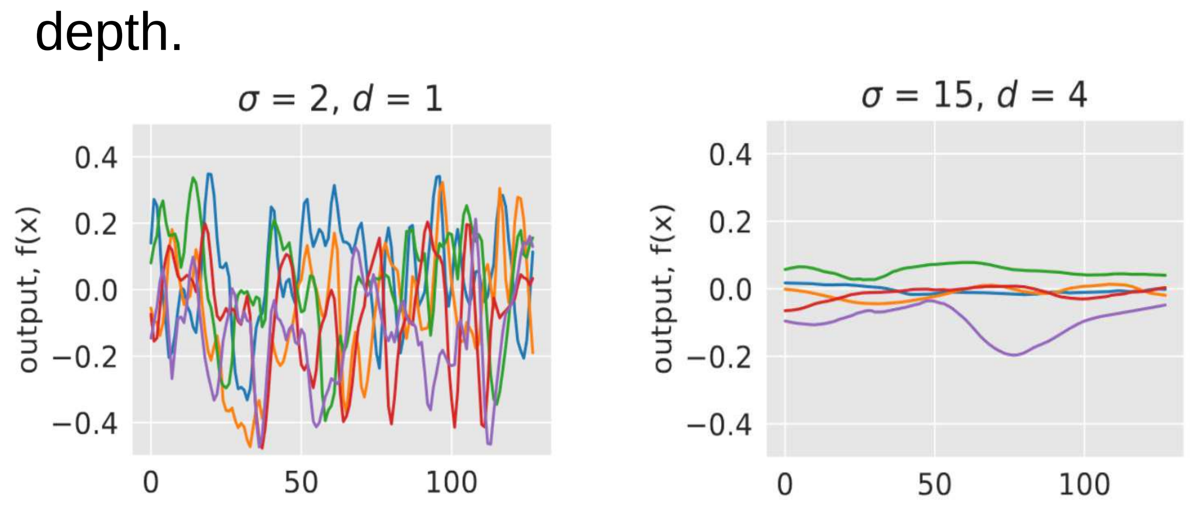
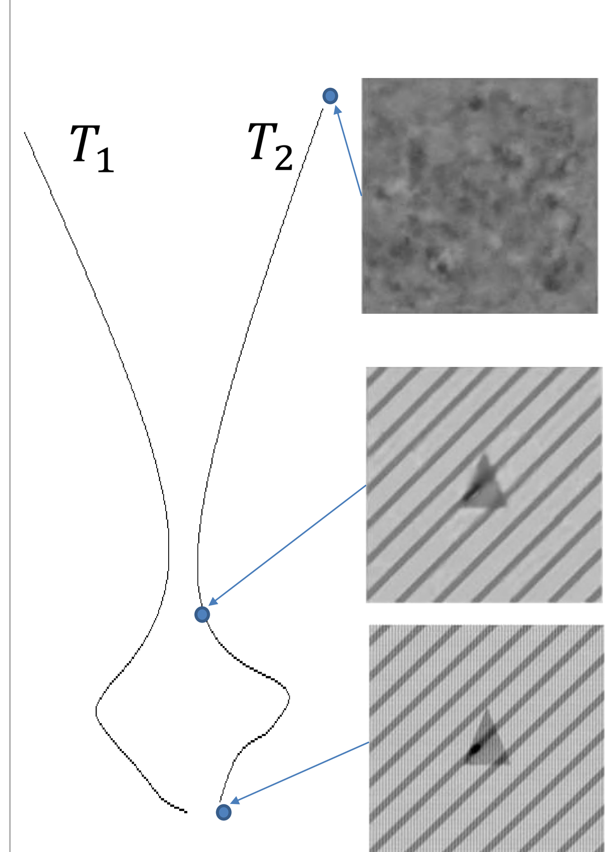


Fig. 3(c) from [2]. Output of the DIP model on 1D signals at initialization with varying input variance σ and depth d

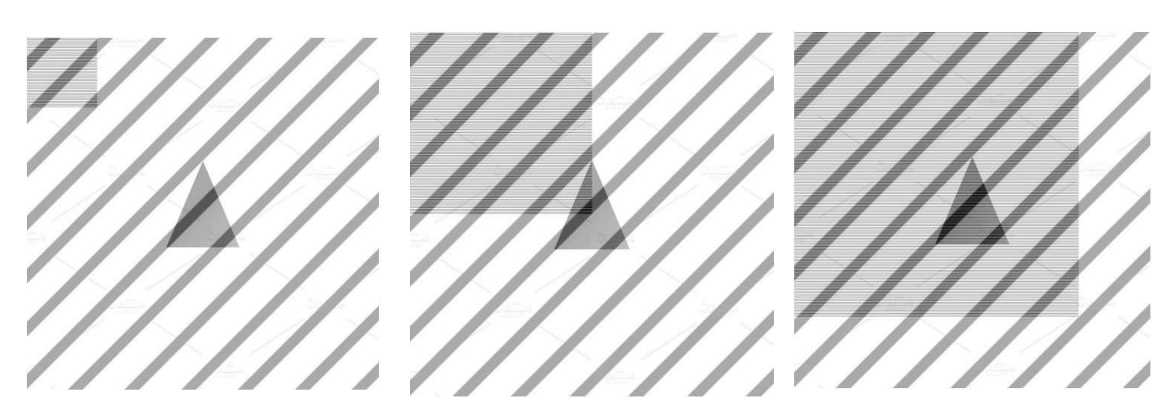
The Spectral Bias

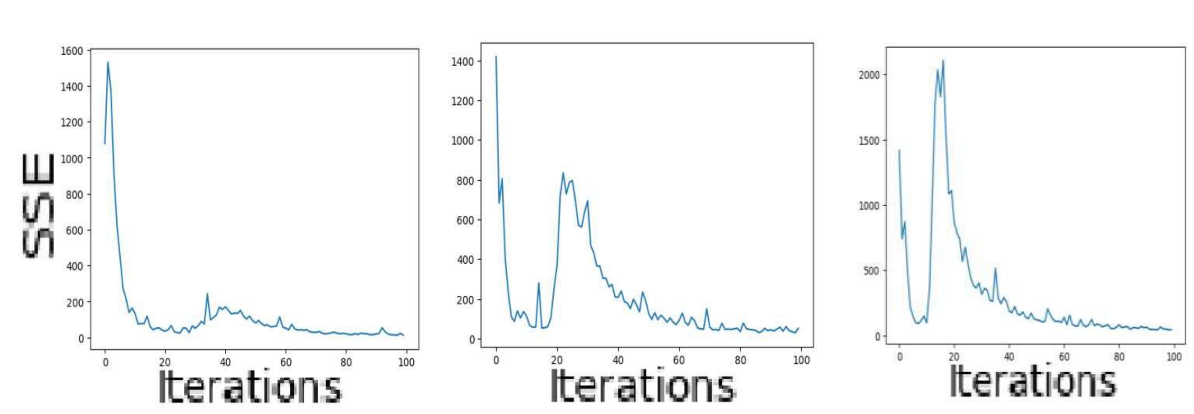
- We study the bias through the entire optimization: DIP learns the frequency components of the image at different rates.
- Low frequency components are learned first.

Demonstrating the bias



- DIP traces out a trajectory in the output space.
- Run the optimization twice. Track the distance between each step of the generated trajectories.
- Trajectories converge twice. The first point of convergence is a low frequency reconstruction of the image.
- Add controlled amounts of high frequency patterns
- Trajectories diverge more after the first convergence as the size of the pattern increases.





The spectral bias causes denoising

- Experiments in [1] denoised additive Gaussian noise.
  This is due to the spectral bias: early stopping makes it behave like a low pass filter.
- Confirming the idea: Use it to generate failure cases.
  Add low frequency noise to images. DIP will fit the noise before the true high frequency patterns in the image.

Barbara.png with low frequency noise | 20 iterations | 500 iterations

What Causes the Frequency Bias?

Convolution layers

- Convolution layers enable frequency decoupling [4].
- Compare models with and without convolutions on image denoising:
  - **DIP**: Model from [1]
  - **DIP Linear-128**: DIP with fully-connected layers, 128 units each
  - **DIP Linear-2048**: DIP Linear, 2048 units each per layer
  - **ReLUNet**: Fully-connected network. 10 layers, 256 nodes each. Maps pixel coordinates to intensities (this will exhibit frequency bias [3]).

Image	DIP	DIP Linear-128	DIP Linear-2048	ReLUNet
Baboon	24.73±0.04	20.48±0.06	20.35±0.43	24.56±0.02
F16	27.00±0.15	20.72±0.12	18.60±1.85	26.67±0.14
House	26.89±0.10	20.69±0.04	18.89±1.45	27.21±0.22
kodim01	25.90±0.08	20.35±0.07	20.54±0.07	25.95±0.07
kodim02	29.20±0.13	20.74±0.01	20.59±0.19	29.99±0.09
kodim03	29.26±0.13	20.47±0.04	18.13±0.13	29.62±0.16
kodim12	29.61±0.06	20.36±0.03	17.74±0.04	29.63±0.16
Lena	27.49±0.06	20.48±0.01	18.67±0.72	27.78±0.16
Peppers	27.12±0.11	20.60±0.01	18.98±1.31	26.82±0.16
	27.47	20.54	19.17	27.58

- ReLUNet and DIP-Linear do not explicitly have convolutions. These are related to special cases of DIP: kernel size 1 and the size of the entire signal.

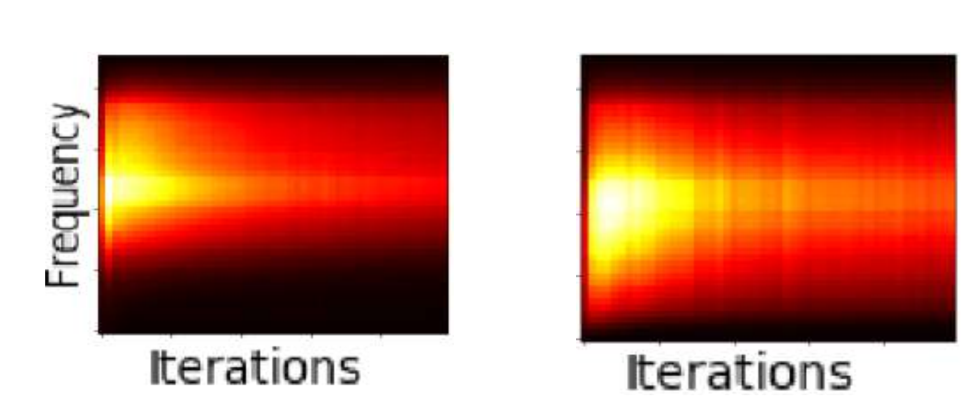
Initialization | 100 iterations | 500 iterations

Initialization | 10000 iterations | 50000 iterations

Outputs with DIP (top row) and ReLUNet (bottom row). ReLUNets are more strongly biased to low frequencies.

Upsampling layers

- Experiments in [1] explored nearest neighbor and bilinear upsampling.
  Both methods are biased towards smooth outputs.
- Low error region in the power spectrum grows slower with stride 4 than 32.
- A larger stride produces a smoother output at the same iteration.



References

1. D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 9446–9454, 2018.
2. Z. Cheng, M. Gadelha, S. Maji, and D. Sheldon. A bayesian perspective on the deep image prior. Computer Vision and Pattern Recognition (CVPR), 2019.
3. N. Rahaman, D. Arpit, A. Baratin, F. Draxler, M. Lin, F. A. Hamprecht, Y. Bengio, and A. Courville. On the spectral bias of deep neural networks. International Conference on Machine Learning (ICML) 2019.
4. A. M. Saxe, P. W. Koh, Z. Chen, M. Bhand, B. Suresh, and A. Y. Ng. On random weights and unsupervised feature learning. In ICML, volume 2, page 6, 2011.