Unlearned Neural Networks as Image Priors for Inverse Problems

Imaging inverse problems

x – image $\mathbb{R}^n$
y – observation $\mathbb{R}^m$
F – forward model $\mathbb{R}^n \to \mathbb{R}^m$
η – noise

$$y = F(x) + \eta$$

Given: F, y
Find: x

Examples

$F = I$ – denoising

$F =$ subset of rows of I — inpainting

$F =$ random $A \in \mathbb{R}^{m \times n}$ w/ $m<n$ — Compressed Sensing

Neural network approaches we've seen

End-to-end

Choose architecture: $f_\theta : \mathbb{R}^m \to \mathbb{R}^n$

Training:
- Collect/create $\mathcal{D} = \{(x_i, y_i)\}$
- Learn θ s.t. $x_i \approx f_\theta(y_i)$

Inversion:
- Given y, compute $x = f_\theta(y)$

Generative Modeling

Choose architecture $g_\theta : \mathbb{R}^k \to \mathbb{R}^n$

Training:
- Collect $\mathcal{D} = \{x_i\}$
- Learn θ s.t. $g_\theta(z)$ for $z \sim \mathcal{N}(0, I_k)$ approximates density from which $\mathcal{D}$ was sampled

Inversion:
- Given y, Find z st $F(g_\theta(z)) \approx y$

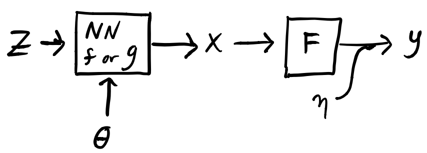
Unlearned Neural Network Priors

Choose architecture $f_\theta : \mathbb{R}^k \to \mathbb{R}^n$

Training: None

Inversion: Fix $z \in \mathbb{R}^k$. Given y,
find θ s.t. $F(f_\theta(z)) \approx y$

Comparison of gen. models and unlearned priors



Gen. Models

- θ – learned from data, general to all images
- z – fit to measurements, specific to an image

Unlearned Priors

- z – fixed, general to all images
- θ – fit to measurements, specific to an image

Can a neural network w/ no training be a good prior?

Yes! The architecture has bias toward some images/signals over others

Deep Image Prior	Ulyanov et al.
Deep Decoder	Heckel and Hand
Deep Geometric Prior	Williams et al.

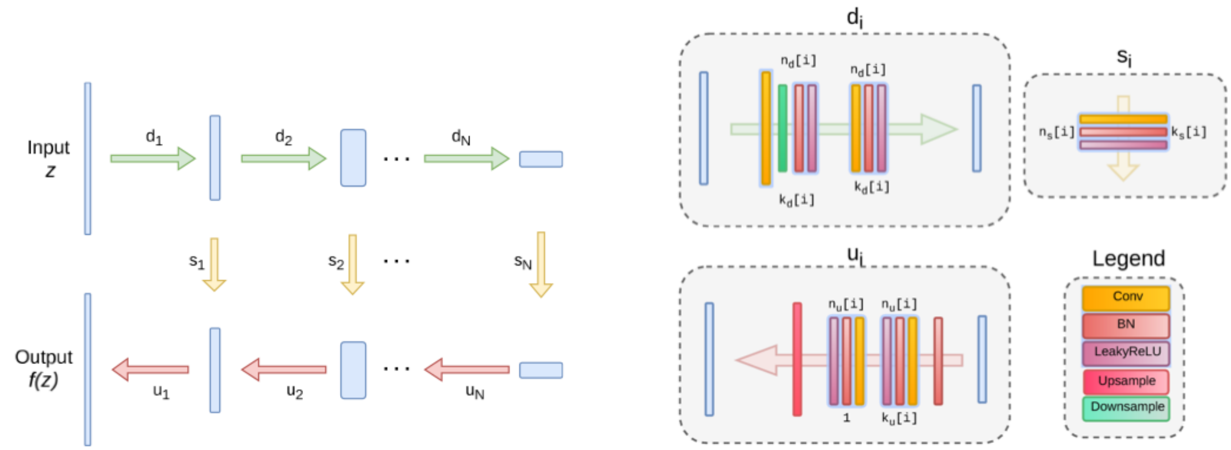
They can also aid learned priors

Image Adaptive GAN	Hussein et al.
Latent Convolutional Models	Athar et al.

Deep Image Prior

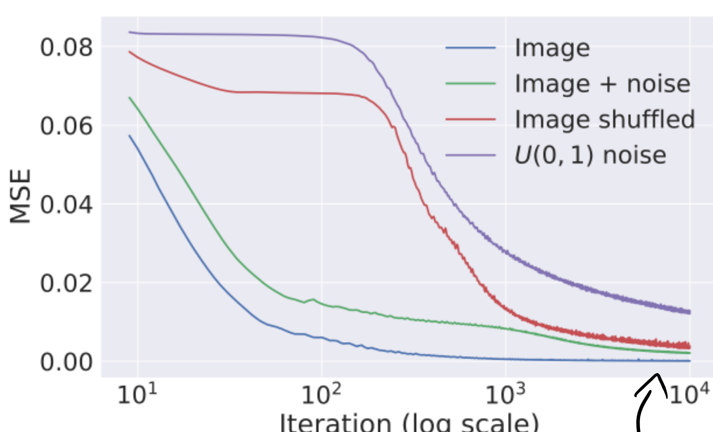
An untrained net whose weights can be optimized to fit measurements

Encoder - decoder architecture (U-net)



DIP can represent all images, but has high impedence to noise and low impedance to signal

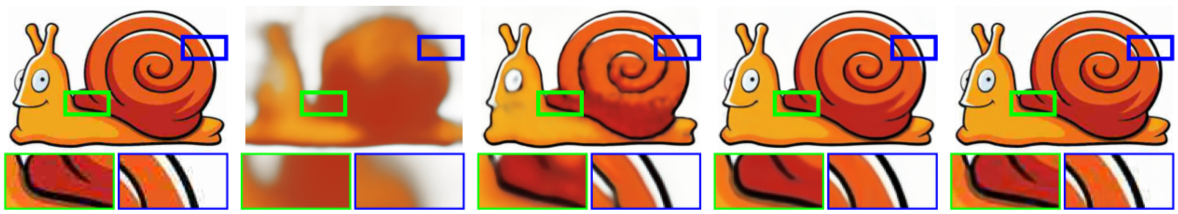
Illustration: Given a single image X, $\min_{\theta} \|f_\theta(z) - X\|$



can fit noise

but doesn't until after the signal is fit

DIP with Early Stopping can denoise

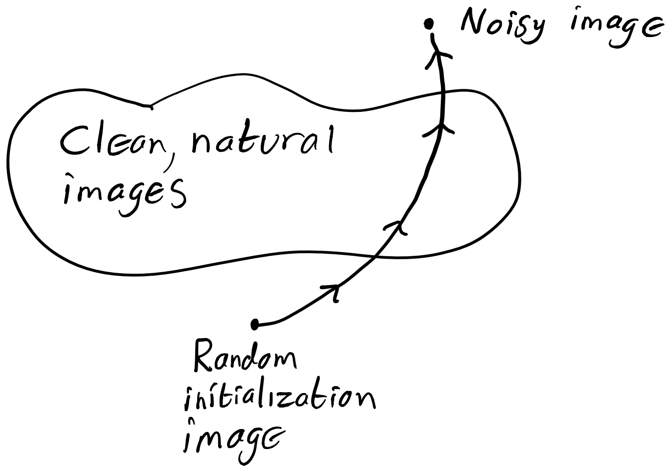


Corrupted | 100 iterations | 600 iterations | 2400 iterations | 50K iterations

Can fit any noisy image...

... but fits a clean version first

Geometric Picture



DIP can also do superresolution, some inpainting

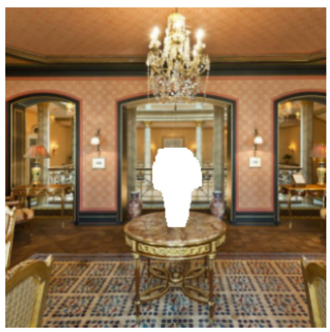
Superresolution

Original

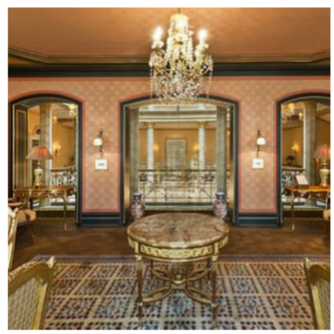
Bicubic Upsampling
4x

DIP
4x

Inpainting



Masked image



DIP

Q: Could DIP inpaint these images well?

Q: For compressed sensing, would DIP need early stopping?

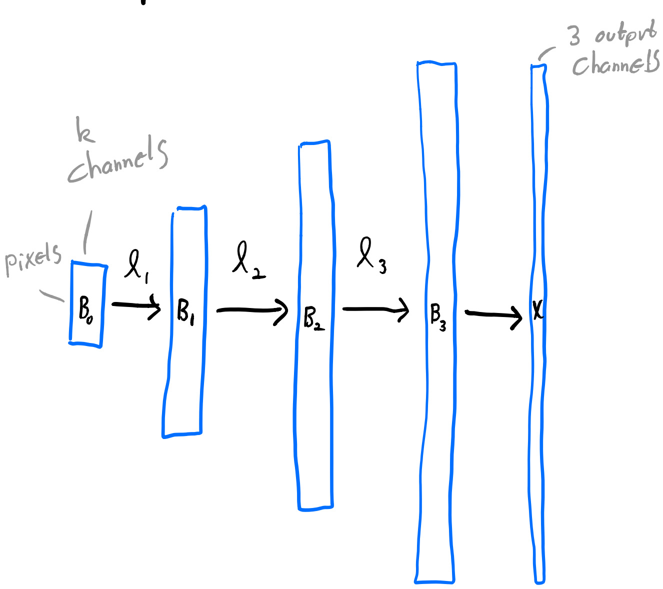
Deep Decoder

Architectural simplification of DIP
- only a decoder

Can be underparameterized (#NN weights < #pixels)
- Concise image representation
- Doesn't need early stopping
- Admits some theory (can't fit noise)

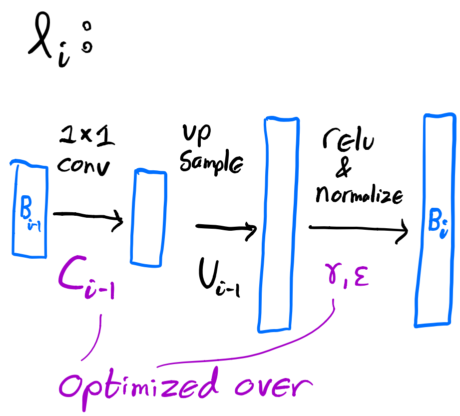
Overparameterized variants can be used for compressed sensing

Deep Decoder Architecture



Typically # channels is kept constant

l_i:



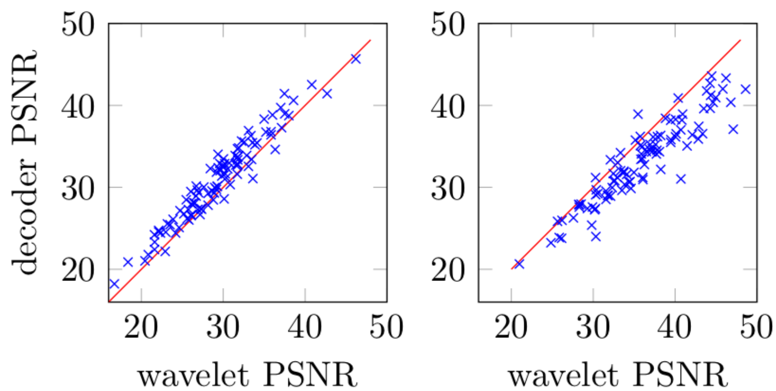
$$B_{i+1} = \underbrace{cn}_{\text{channelwise normalization}}(relu(\underbrace{U_i}_{\text{fixed}} B_i C_i))$$

Parameters of deep decoder: $\Theta = \{ \underbrace{C_0, \dots C_{d-1}}_{\mathbb{R}^{k \times k}}, \underbrace{C_d}_{\mathbb{R}^{k \times 3}} \} \cup \{ \text{normalization params} \}$

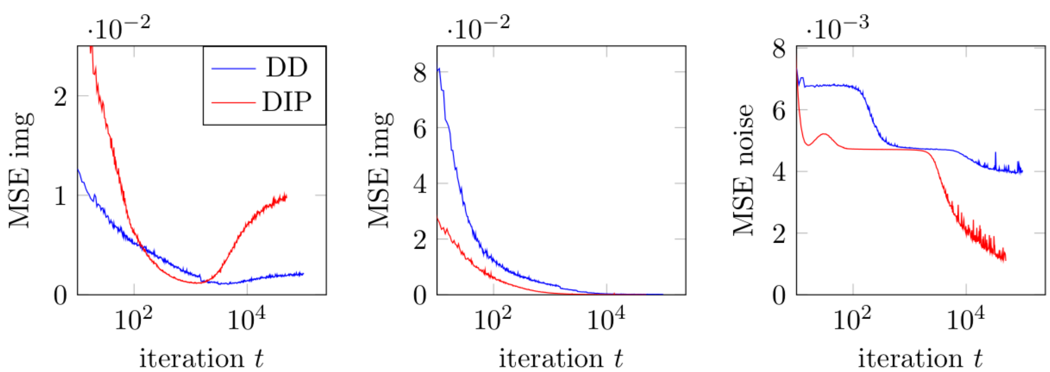
Roughly $\approx k^2 d$ params

Deep Decoder can concisely represent images

Given image x, $\min_{\theta} \| f_\theta(z) - x \|$



Underparameterized Deep Decoder can denoise without early stopping



(a) fit noisy image (b) fit image (c) fit noise

Theory: Underparameterized deep decoder
Can't fit noise

Claim: Consider a 1-layer DD. Fix $B_0 \in \mathbb{R}^{n_0 \times k}$, $U_0 \in \mathbb{R}^{n \times n_0}$

$$G(C_0, C_1) = \mathrm{relu}(\underbrace{U_0 B_0}_{\text{fixed, arbitrary}} C_0) C_1$$

Let $\eta \sim \mathcal{N}(0, \sigma^2 I_n)$

If $\frac{k^2 \log n_0}{n} \leq \frac{1}{32}$, then with high probability

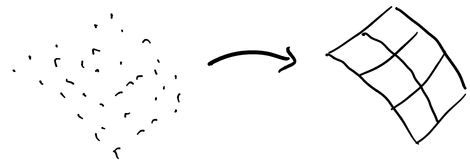
$$\min_{C_0, C_1} \| G(C_0, C_1) - \eta \|^2 \geq \|\eta\|^2 \left(1 - 20 \frac{k^2 \log n_0}{n}\right)$$

Rough Rationale: This DD has $\approx k^2$ params.
A k^2-dim subspace of $\mathbb{R}^n$ will have $\frac{k^2}{n}$ fraction of the noise power.

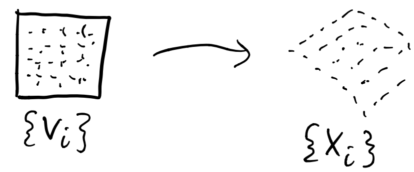
Proof idea: The range of G lives in union of at most $n_0^{k^2}$ different k^2-dim subspaces.

Deep Geometric Prior

Given noisy points $\{x_i\}$ along a surface in $\mathbb{R}^3$, estimate the surface



A local patch can be represented as
$\phi : [0,1] \times [0,1] \to \mathbb{R}^3$



Estimate Surface:

- Choose some $\{V_i\} \in [0,1] \times [0,1]$
- Find best matching $\{V_i\} \leftrightarrow \{X_i\}$ by minimizing Earth Mover Distance
- Represent ϕ by a fully connected ReLU-net

$$\phi(V;\theta) = \theta_d \, \text{relu}(\theta_{d-1} \, \text{relu}(\cdots \text{relu}(\theta_1 v) \cdots))$$

w/ θ_i matrices

- Solve

$$\min_{\theta} \sum_i \| \phi(V_i;\theta) - X_i \|^2$$

During optimization, overall shape gets fit first, then eventually the noise. So, early stop.

Can also apply to images:

Image X can be viewed as a map $[0,1] \times [0,1] \to \mathbb{R}^3$

Estimate this map by a net

position

3 color channels

What is the prior w/ an unlearned net?

Underparameterized Deep Decoder

- $\{ f_\theta(z) \mid \theta \}$ for fixed z is a low-dim manifold in $\mathbb{R}^n$.
- The prior is membership in this set.
- can only fit low-complexity signals (don't have a characterization)

Overparameterized Priors

- Can fit anything
- "prefers" low complexity signals
  - piecewise smooth? small Lipschitz const?
  - gradient descent traverses these signals first

Where is smoothness/locality with images enforced?

DIP — upsampling, convolutions, sometimes input

DD — upsampling

DGP — continuous function applied to continuous input

Bringing ideas of unlearned nets to learned nets

- IAGAN – use a trained GAN as a warm start for a DIP

$$\begin{tikzcd} z \arrow[r] & \boxed{G} \arrow[r] & x \\ & \theta \arrow[u] & \end{tikzcd}$$

- optimize z & θ in an image specific way at inversion.

$$\min_{z,\theta} \| F(G_\theta(z)) - y \|^2$$

- Latent Convolutional Models
  - use DIP as image specific prior for high dimensional latent representations for a trained generator