Unsupervised Learning with Stein's Unbiased Risk Estimator: With Applications to Denoising and Compressive Sensing

Chris Metzler, Ali Mousavi, Reinhard Heckel, Rich Baraniuk

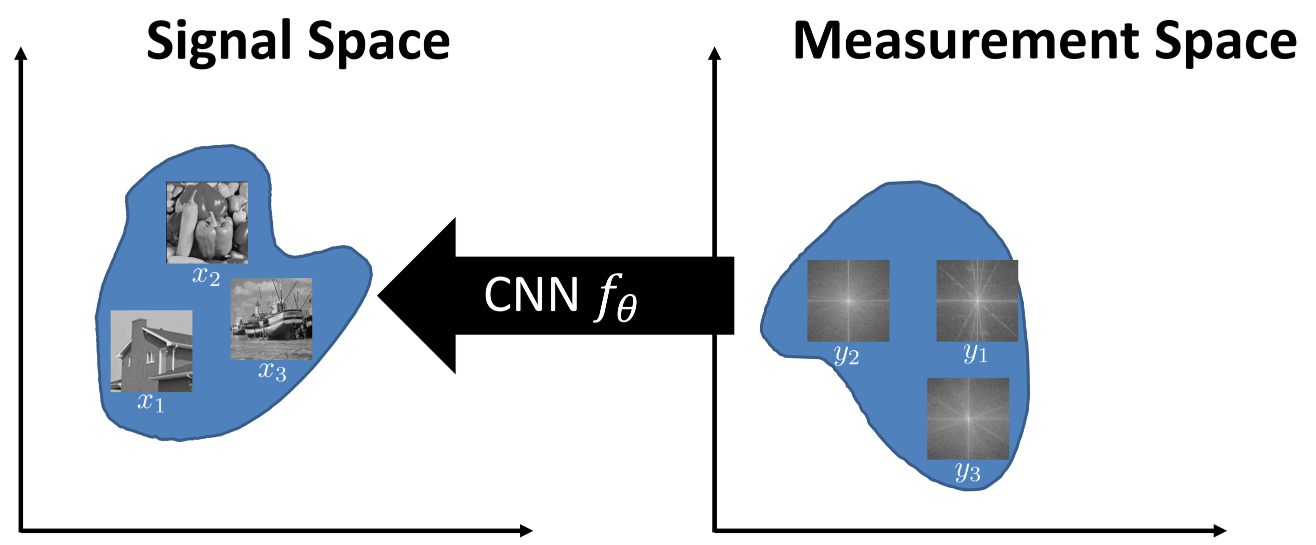
Paper and Code:
https://arxiv.org/abs/1805.10531
https://github.com/ricedsp/D-AMP_Toolbox

CNNs for Imaging Inverse Problems

Many imaging systems, including cameras, MRIs, sonar, radar, and more, capture noisy linear measurements $y_i \in \mathbb{R}^m$ of a scene $x_i \in \mathbb{R}^n$:

$$y_i = \boldsymbol{A}\mathrm{x_i} + \mathrm{w_i}.$$

One must then reconstruct the scene x_i from the measurement y_i.
State-of-the-art: Use a convolutional neural net (CNN), $f_\theta(y_i)$, to learn a mapping $y \mapsto x$.

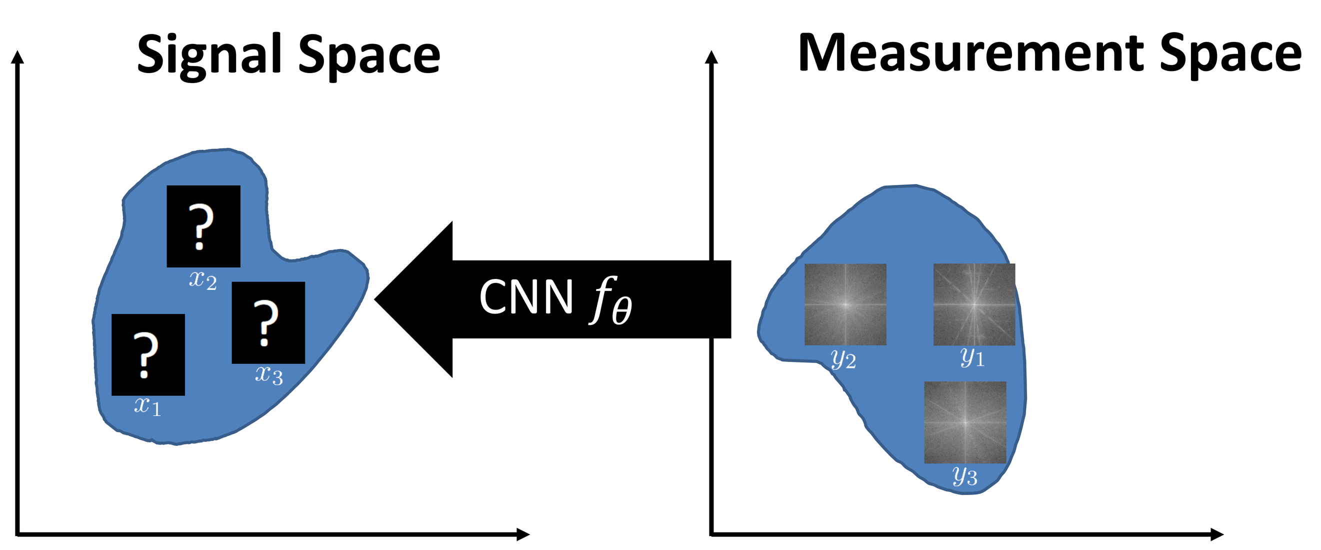


CNN is typically trained by minimizing the MSE:

$$\theta^* = \mathrm{argmin}_\theta \sum_i ||x_i - f_\theta(y_i)||^2.$$

Problem: Need Training Pairs to Compute MSE

In many applications, e.g., medical imaging, we have access to measurements $y_1, y_2, y_3, \ldots$ but never have access ground truth scenes $x_1, x_2, x_3, \ldots$.



Without access to (x_i, y_i) training pairs, we cannot minimize the MSE:

$$\theta^* = \mathrm{argmin}_\theta \sum_i ||? - f_\theta(y_i)||^2.$$

Solution: Stein's Unbiased Risk Estimator

Stein's Unbiased Risk Estimator (SURE) [Stein '81] and its generalizations [Eldar '09, Luisier et al. '11] let you estimate the MSE between an estimate $\hat{x}_i = f(y_i)$ and a scene x_i without accessing x_i. I.e.,

$$\mathbb{E}_w[||x - f_\theta(y)||^2] = \mathbb{E}_w[||y - f_\theta(y)||^2] - n\sigma_w^2 + 2\sigma_w^2 \mathrm{div}_y(f_\theta(y)),$$

where $\mathrm{div}_y(f_\theta(y)) = \sum_n \partial f_{\theta_n}(y)/\partial y_n$.

When the divergence cannot be computed analytically, we can compute it using a Monte Carlo procedure [Ramani et al. '08]:

$$\mathrm{div}_y(f_\theta(y)) \approx b^t \left(\frac{f_\theta(y+\epsilon b) - f_\theta(y)}{\epsilon}\right).$$

We propose training CNNs using the SURE estimate of the MSE.

Denoising Using CNNs Without Training Data

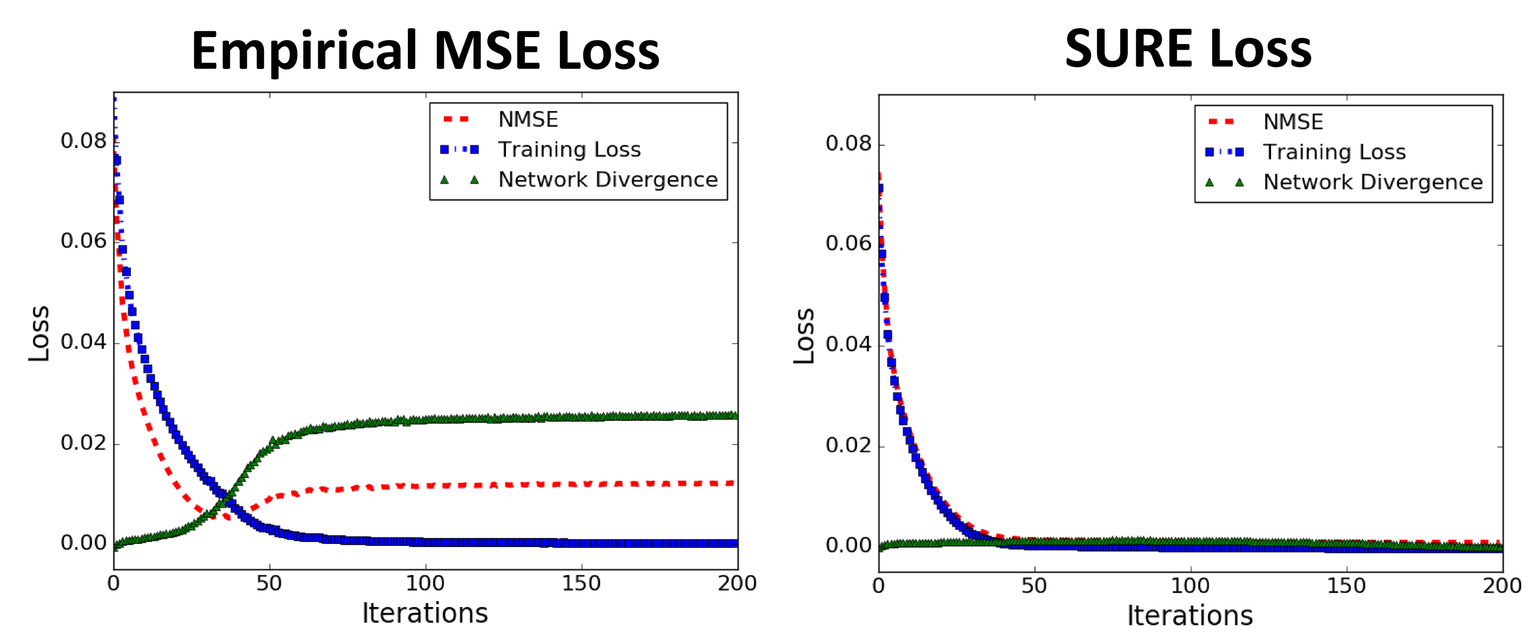
CNNs can be "trained" to denoise a noisy image using just that image.

Setup:

- U-net f_θ initialized with random weights.
- One noisy observation $\mathrm{y} = x + w$, with $w \sim N(0, \sigma^2\mathbf{I})$ with σ^2 known.

Prior Work: [Deep Image Prior by Ulyanov et al. 2017] Given only one noisy measurement, minimize the *empirical* MSE, $||y - f_\theta(y)||^2$.

Proposed: Using only one noisy measurement and prior knowledge of σ^2, minimize the SURE estimate of the *true* MSE, $||x - f_\theta(y)||^2$.



CBM3D: 25.9 dB, 4.84 sec
DnCNN: 26.4 dB, 0.01 sec
SURE U-net: 26.3 dB, 72.15 sec

Resulting network effectively denoises the given image but does not generalize and "retraining" for every new image is impractically slow.

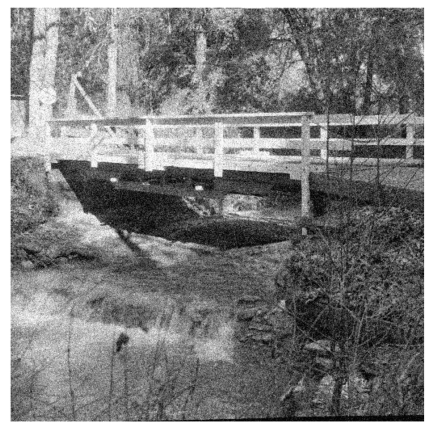
Training CNNs to Denoise Using Only Noisy Data

CNNs can be trained to denoise images using only noisy observations.

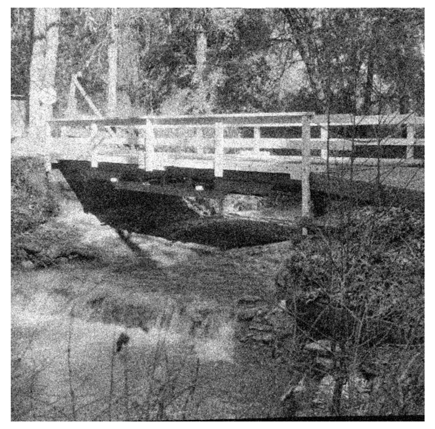
Setup:

- DnCNN f_θ initialized with random weights.
- Thousands of noisy observation $y_i = x_i + w_i$ with $w_i \sim N(0, \sigma^2\mathbf{I})$ with σ^2 known.

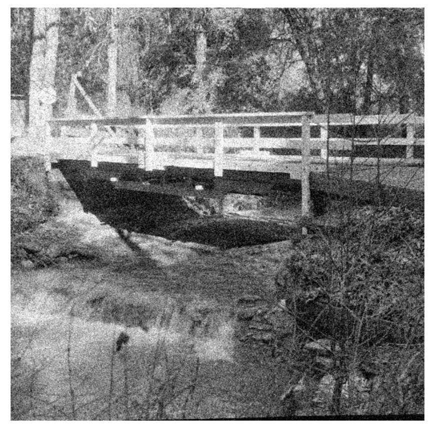
Proposed: Using σ and the many noisy observations, minimize the SURE estimate of the MSE. (Concurrently proposed by [Soltanayev & Chun '18]).



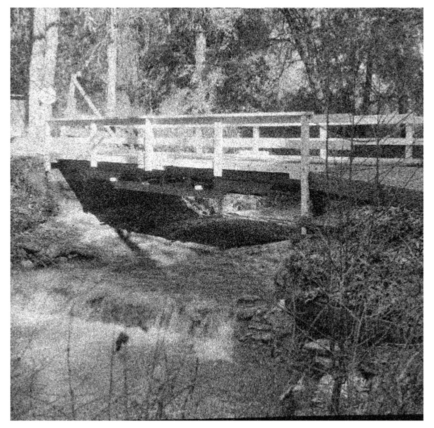
Original Noisy Image



BM3D
(26.0 dB, 4.01 sec)



DnCNN (MSE)
(26.7 dB, 0.04 sec)



DnCNN (SURE)
(26.5 dB, 0.04 sec)

The resulting network offers near state-of-the-art performance.

Training CNNs to Perform Compressive Sensing Recovery Using Only Noisy Measurements

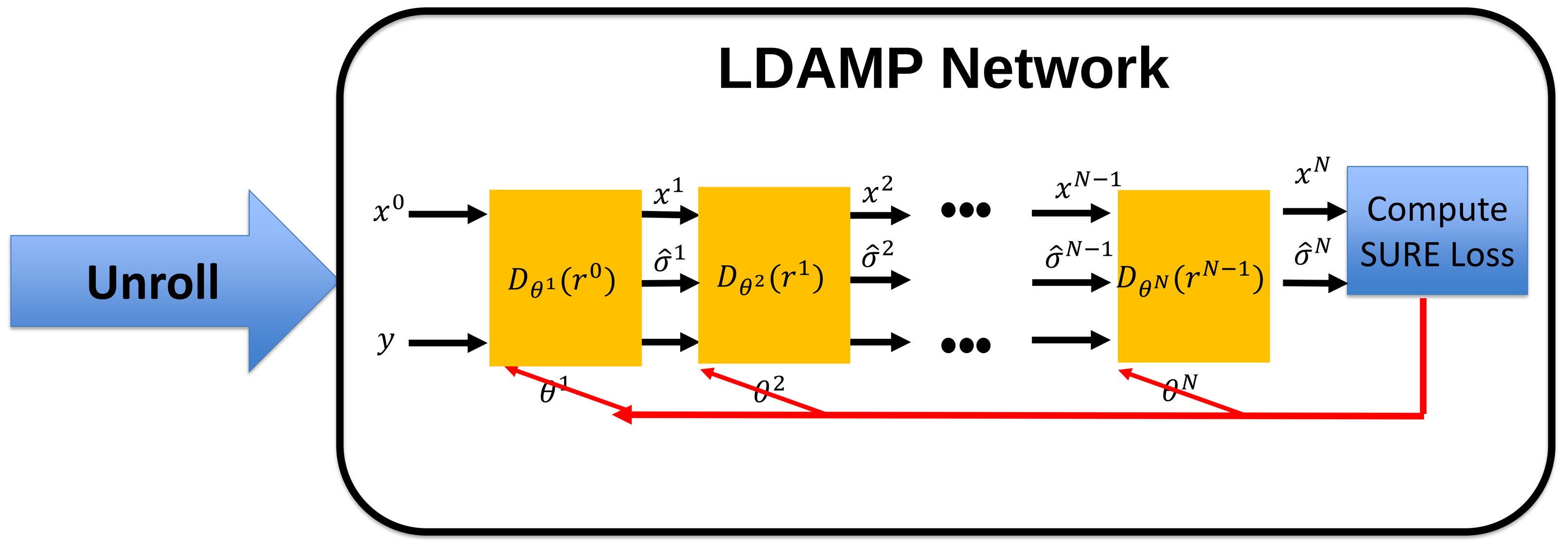
CNNs can be trained to perform compressive sensing recovery using only noisy measurements $y_i = \boldsymbol{A}\mathrm{x_i} + \mathrm{w_i}$.

Background:

- The LDAMP network is formed by unrolling [LeCun et al. '10] the D-AMP [Montanari et al. '09, Metzler et al. '14] algorithm and replacing the typical nonlinearity/denoiser with a CNN.
- AMP algorithms have the property that at every iteration, r^t resembles $x_o + v^t$ where $v^t \sim N(0, {\sigma^t}^2\mathbf{I})$ with σ^t known [Montanari '12].

D-AMP Algorithm

$$\begin{aligned} b^t &= \frac{z^{t-1}\mathrm{div} D_{\theta^{t-1}}(x^{t-1} + \mathbf{A}^H z^{t-1})}{m}, \\ z^t &= y - \mathbf{A}x^t + b^t, \\ \hat{\sigma}^t &= \frac{\|z^t\|_2}{\sqrt{m}}, \\ r^t &= x^t + \mathbf{A}^H z^t, \\ x^{t+1} &= D_{\theta^t}(r^t). \end{aligned}$$



Setup:

- LDAMP network f_θ initialized with random weights.
- Thousands of noisy observation $y_i = \boldsymbol{A}_i x_i + w_i$ with $\boldsymbol{A}_i$ (approximately) i.i.d. Gaussian and known and $w_i \sim N(0, \sigma^2\mathbf{I})$ with σ^2 unknown.

Proposed: Using only noisy compressive measurements and AMP's estimates of σ^t, minimize the SURE estimate of the MSE.

Compressive Reconstructions (20% Sampling):
BM3D-AMP 25.6 dB in 13.2 sec, LDAMP (MSE) 28.6 dB in .4 sec, LDAMP (SURE) 28.0 dB in .4 sec

Take-away: Could throw a compressive sensor into a new environment and let it capture measurements and train itself using SURE.