Solving linear inverse problems using the prior implicit in a denoiser

Zahra Kadkhodaie, Eero P. Simoncelli • New York University

Abstract

Prior probability models are a central component of many image processing problems, but density estimation is notoriously difficult for high-dimensional signals such as photographic images. Deep neural networks have provided state-of-the-art solutions for problems such as denoising, which implicitly rely on a prior probability model of natural images. Here, we develop a robust and general methodology for making use of this implicit prior. We rely on a little-known statistical result due to Miyasawa (1961), who showed that the least-squares solution for removing additive Gaussian noise can be written directly in terms of the gradient of the log of the image signal density. We use this fact to develop a stochastic coarse-to-fine gradient ascent procedure for drawing high-probability samples from the implicit prior embedded within a CNN trained to perform blind (i.e., unknown noise level) least-squares denoising. A generalization of this algorithm to constrained sampling provides a method for using the implicit prior to solve any linear inverse problem, with no additional training. We demonstrate this general form of transfer learning in multiple applications, using the same algorithm to produce high-quality solutions for deblurring, super-resolution, inpainting, and compressive sensing.

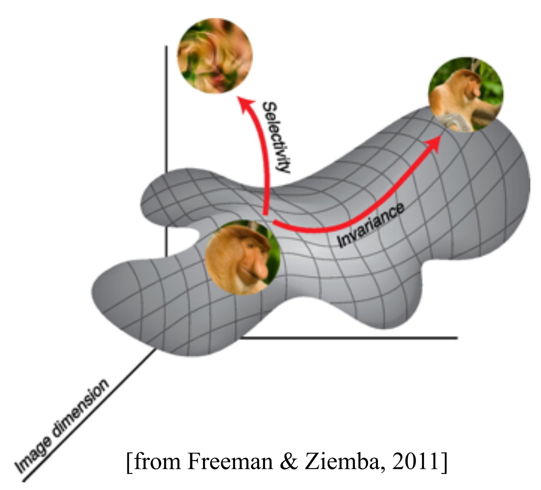
Image priors, manifolds, and noisy observations

Visual images lie on a low-dimensional manifold, spanned by various natural deformations.

Images on this manifold are approximately equally probable - at least locally. Probability of x being a natural image, p(x), is zero everywhere except for x drawn from the manifold.

An observed image, y, contaminated with Gaussian noise, $z \sim N(0, \sigma^2)$ is drawn from an observation density, p(y), which is a Gaussian-blurred version of the image prior.

Moreover, the family of observation densities over different noise variances, $p_\sigma(y)$, forms a Gaussian scale-space representation of the prior analogous to the temporal evolution of a diffusion process.



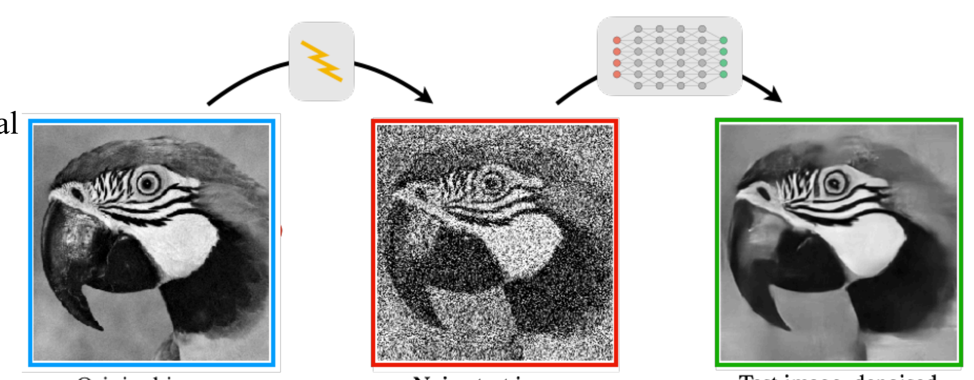
[from Freeman & Ziemba, 2011]

Least squares denoising

Noisy observation: $\boxed{y} = \boxed{x} + z$

The least squares estimate of the true signal is the conditional mean of the posterior:

$$\boxed{\hat{x}(y)} = \min_{\hat{x}} \int \|\hat{x} - x\|^2 \, p(x|y) \, dx$$
$$= \int x \, p(x|y) \, dx$$
$$= \int x \, \frac{p(y|x)p(x)}{p(y)} \, dx$$

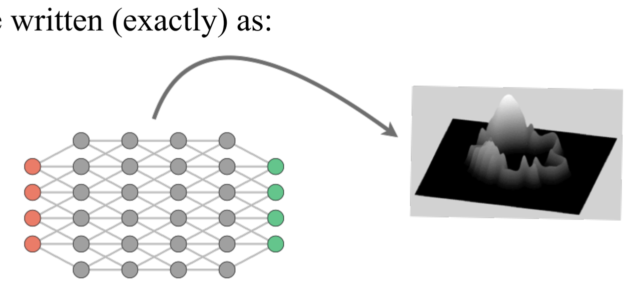


Original image x — Noisy test image, $\sigma = 90$ — Test image, denoised by BF-CNN

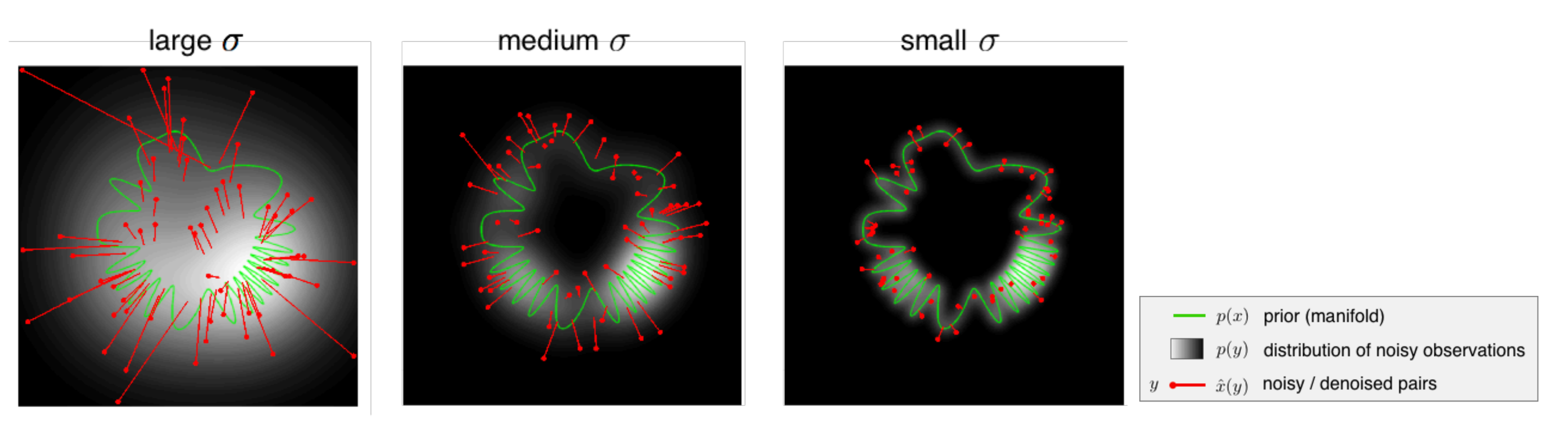
Exposing the implicit prior through Empirical Bayes estimation

For Gaussian noise contamination, the least squares estimate may be written (exactly) as:

$$\hat{x}(y) = \int x \, p(x|y) \, dx$$
$$= y + \sigma^2 \nabla_y \log p(y)$$



This is Miyasawa's Empirical Bayes formulation (1961), which expresses the denoising operation in terms of the gradient of the prior predictive density, *p(y)*.



Two-dimensional simulation/visualization:
End of red line segments shows the least-squares optimal denoising solution $\hat{x}(y)$ for each noisy signal, y.

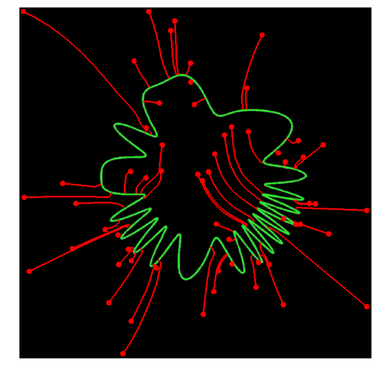
Drawing high-probability samples from the implicit prior

Algorithm in a nutshell:

- Use denoiser-defined gradient to go uphill in probability
- Do this iteratively
- On each step, effect noise decreases, and effective prior becomes less blurred. Gradient step size automatically adapts to each noise level.
- This coarse to fine optimization procedure converges to a point on the manifold!

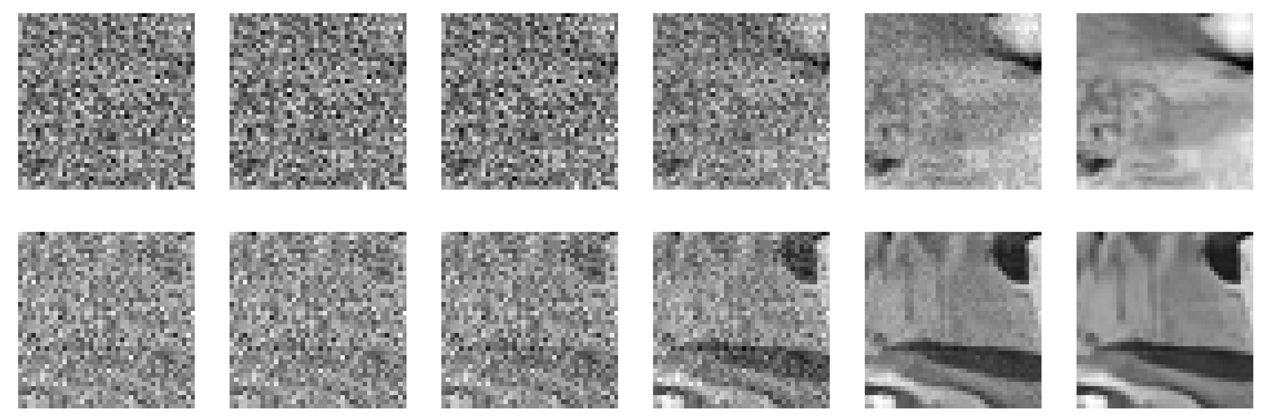
Algorithm 1: Coarse-to-fine stochastic ascent method for sampling from the implicit prior of a denoiser, using denoiser residual $f(y) = \hat{x}(y) - y$.

parameters: $\sigma_0, \sigma_L, h_0, \beta$
initialization: $t = 1$, draw $y_0 \sim \mathcal{N}(0.5, \sigma_0^2 I)$
while $\sigma_{t-1} \leq \sigma_L$ **do**
  $h_t = \frac{h_0 t}{1+h_0(t-1)}$;
  $d_t = f(y_{t-1})$;
  $\sigma_t^2 = \frac{\|d_t\|^2}{N}$;
  $\gamma_t^2 = \left((1-\beta h_t)^2 - (1-h_t)^2\right)\sigma_t^2$;
  Draw $z_t \sim \mathcal{N}(0, I)$;
  $y_t \leftarrow y_{t-1} + h_t d_t + \gamma_t z_t$;
  $t \leftarrow t+1$
end

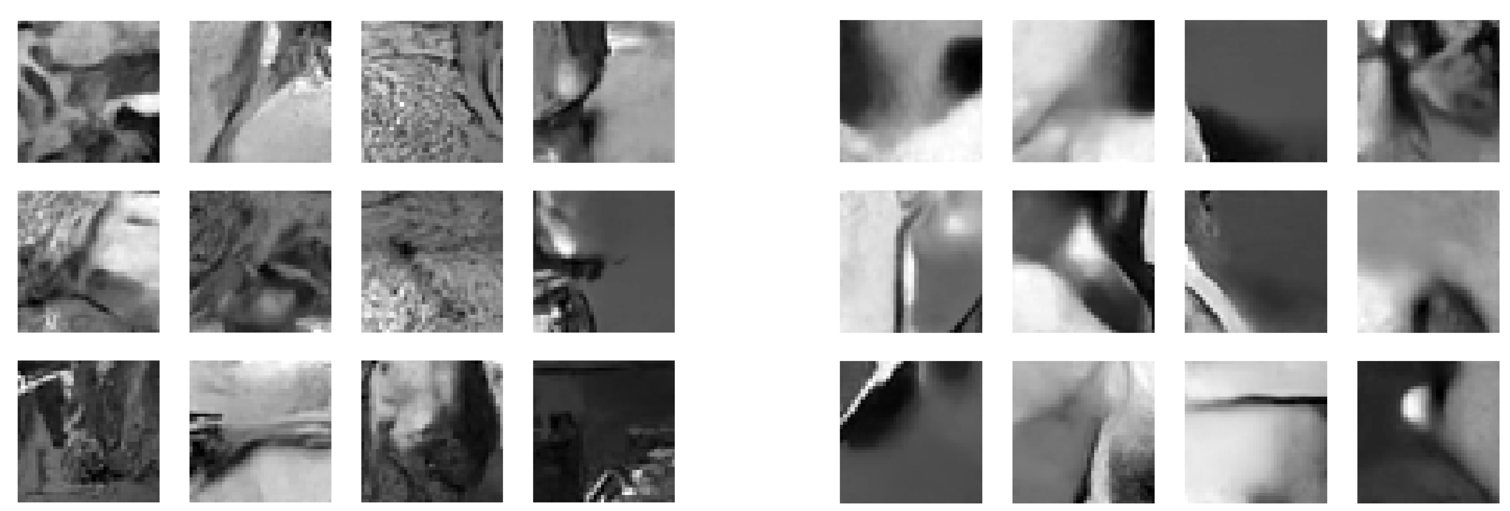


Two-dimensional visualization: trajectory of our iterative coarse-to-fine inverse algorithm

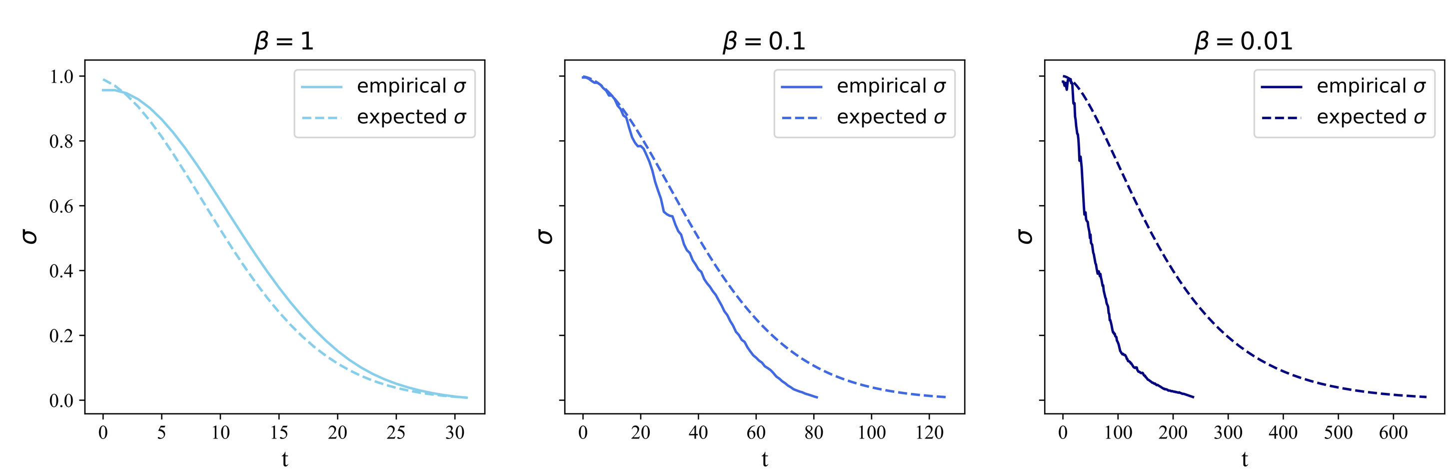
Schedule for noise amplitude, γ_t, ensures that noise drops as $\sigma_t^2 = (1 - \beta h)^2 \sigma_{t-1}^2$, where $\beta \in [0, 1]$



Two sequences of images, y_t, from the iterative sampling procedure, with different initializations, y_0, and no added noise ($\beta = 1$).



More samples arising from different inializations. Left: A moderate level of noise ($\beta = 0.5$) is injected in each iteration. Right: A high level of injected noise ($\beta = 0.1$).



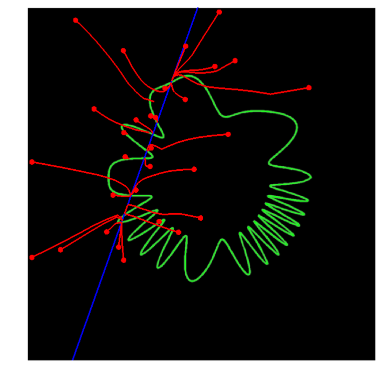
Convergence of σ_t for three synthesized patches with different values of β. Convergence is faster than the scheduled rate.

Solving linear inverse problems using the implicit prior

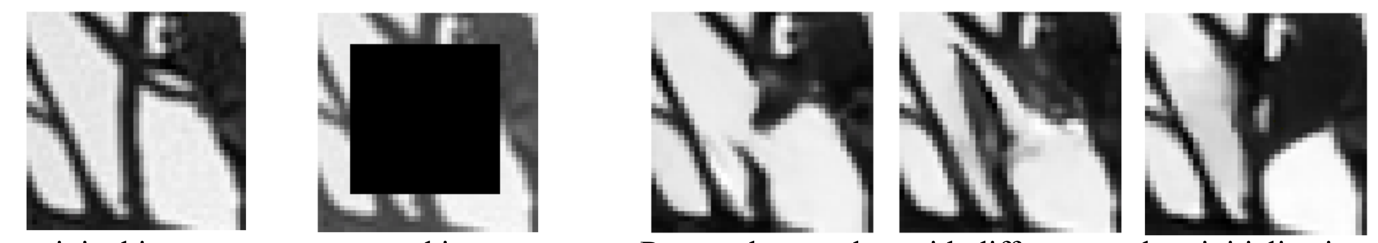
Given a set of linear measurements of an image, $x^c = M^T x$, where M is a low-rank measurement matrix, we use an enhanced version of our algorithm to recover the original image

Algorithm 2: Coarse-to-fine stochastic ascent method for sampling from $p(x|M^T x = x^c)$, based on the residual of a denoiser, $f(y) = \hat{x}(y) - y$. Note: e is an image of ones.

parameters: $\sigma_0, \sigma_L, h_0, \beta, M, x^c$
initialization: t=1; draw $y_0 \sim \mathcal{N}(0.5(I - MM^T)e + Mx^c, \sigma_0^2 I)$
while $\sigma_{t-1} \leq \sigma_L$ **do**
  $h_t = \frac{h_0 t}{1+h_0(t-1)}$;
  $\boxed{d_t = (I - MM^T)f(y_{t-1}) + (Mx^c - MM^T y_{t-1});}$
  $\sigma_t^2 = \frac{\|d_t\|^2}{N}$;
  $\gamma_t^2 = \left((1-\beta h_t)^2 - (1-h_t)^2\right)\sigma_t^2$;
  Draw $z_t \sim \mathcal{N}(0, I)$;
  $y_t \leftarrow y_{t-1} + h_t d_t + \gamma_t z_t$;
  $t \leftarrow t+1$
end



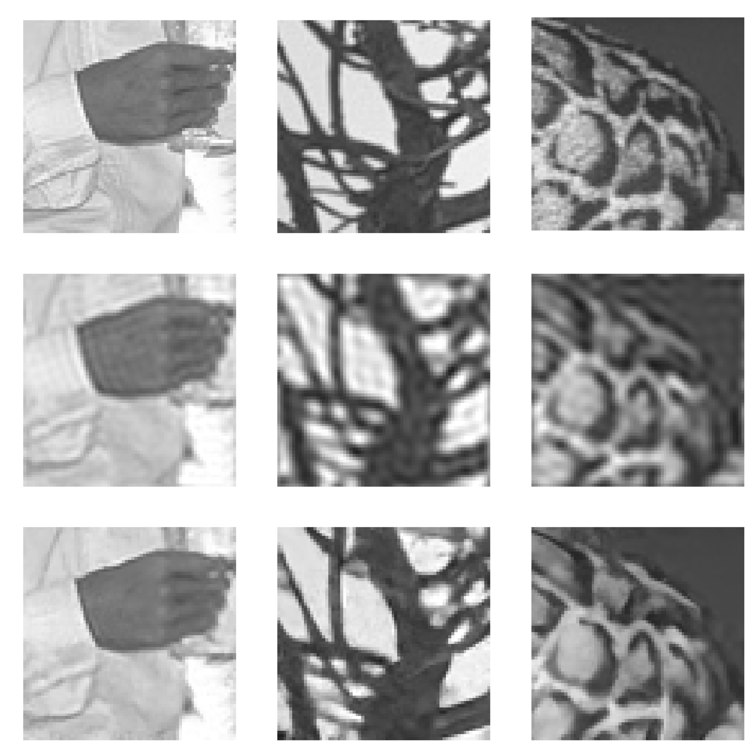
Inpainting



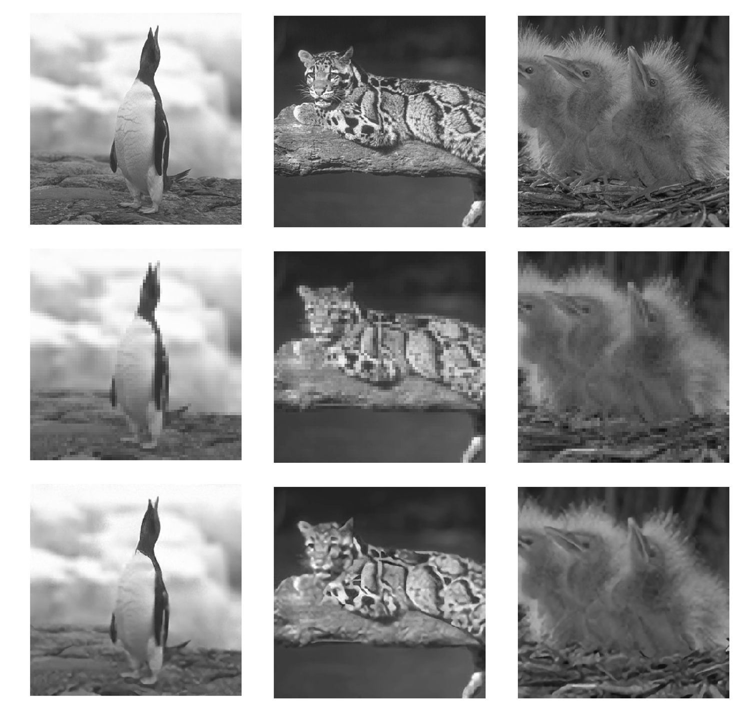
original image — corrupted image — Restored examples, with different random initializations

In all the following four linear inverse applications, original image is in the top row, corrupted image is in the middle row, and restored image is in the bottom row.

De-blurring
10% of frequencies preserved.

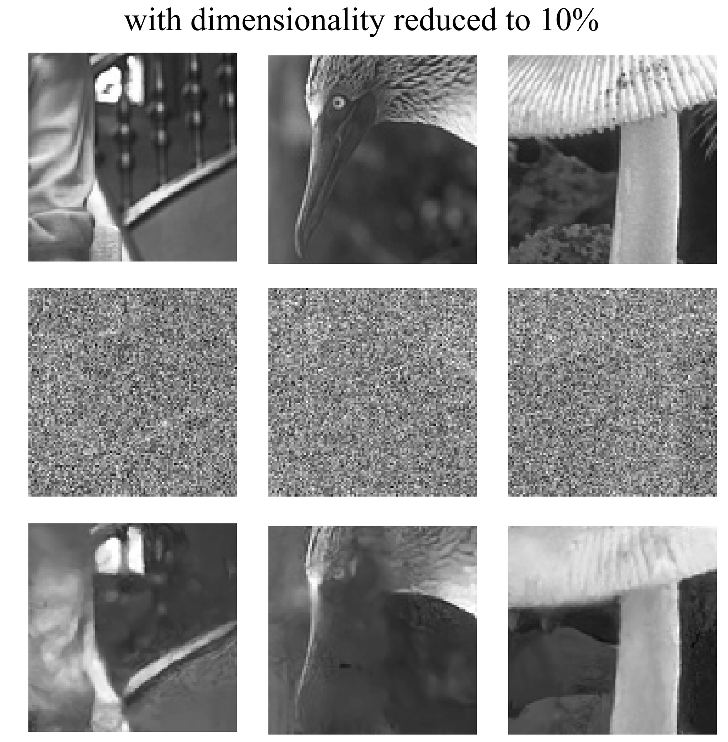


Super-resolution
Resolution reduced by averaging over 4x4 blocks (dimensionality reduction to 6.25%).



Random missing pixels
10% of pixels prserved.

Compressive sensing
Measurement matrix M contains random, orthogonal unit vectors, with dimensionality reduced to 10%



Further information: ArXiv 2007.13640, Jul 2020