DeCompress: Denoising via Neural Compression

Learn to Compress & Compress to Learn Workshop, 2025

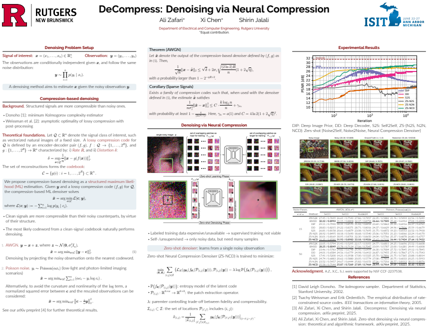
Ali Zafari* Xi Chen* Shirin Jalali

Department of Electrical and Computer Engineering
Rutgers University

Pre-presentation: Beyond DeCompress

- Extended algorithm and theoretical insights:
  - *Zero-shot Denoising via Neural Compression*: arXiv:2506.12693

- Happy to chat - stop by our poster ⇓



Outline

Denoising Problem

- Signal of interest: $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}_+^n$

- Observations: $\boldsymbol{y} = (y_1, \dots, y_n)$
  - conditionally independent given $\boldsymbol{x}$

$$\boldsymbol{y} \sim \prod_{i=1}^{n} p(y_i \mid x_i).$$

Denoising Problem

- Signal of interest: $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$

- Observations: $\boldsymbol{y} = (y_1, \ldots, y_n)$
  - conditionally independent given $\boldsymbol{x}$

$$\boldsymbol{y} \sim \prod_{i=1}^{n} p(y_i \mid x_i).$$

Denoising problem: Estimate $\boldsymbol{x}$ from noisy observation $\boldsymbol{y}$.

- Key ingredient: Using the source structure

Denoising: An old problem, new challenges

- Denoising has a long history with a wide range of approaches
- Mature field under standard setting: Gaussian noise in natural images is well-understood
- Emerging imaging applications introduce non-traditional sources and non-standard noise models, requiring fresh solutions

NLM [Buades et al. 2005], BM3D [Dabov et al. 2007], DnCNN [Zhang et al. 2016], Restormer [Zamir et al. 2022], DIP [Ulyanov et al. 2017], Deep Decoder [Heckel et al. 2018], Noise2Void [Krull et al. 2019], Noise2Score [Kim et al. 2021], ...

Compression-based Denoising

- Clean signal $\boldsymbol{x}$ is structured, and *compressible*
- Compression-based denoising idea:

  Lossy compression of noisy signal at a distortion level adjusted based on noise power $\equiv$ denoising

Compression-based Denoising

- Clean signal $\boldsymbol{x}$ is structured, and *compressible*
- Compression-based denoising idea:

  Lossy compression of noisy signal at a distortion level adjusted based on noise power $\equiv$ denoising

- Key theoretical works on compression-based denoising:
  - Minimum Kolmogorov Complexity Estimator [Donoho, 2002]
  - Universal denoising of discrete-valued sources [Weissman et al., 2005]: universal lossy compression followed by suitable post-processing

Compression-based Denoising

- Clean signal $\boldsymbol{x}$ is structured, and *compressible*
- Compression-based denoising idea:

  Lossy compression of noisy signal at a distortion level adjusted based on noise power $\equiv$ denoising

- Key theoretical works on compression-based denoising:
  - Minimum Kolmogorov Complexity Estimator [Donoho, 2002]
  - Universal denoising of discrete-valued sources [Weissman et al., 2005]: universal lossy compression followed by suitable post-processing
- Despite strong theoretical foundations, existing compression-based denoising methods i) often yield sub-optimal results in practice, 2) require access to pre-defined compression codes.

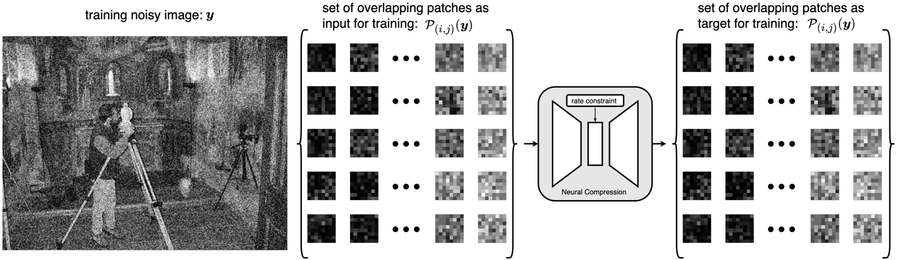
Our contribution

- Inspired by the recent success of neural compression methods, we introduce DeCompress:
  - a novel compression-based denoising method
- Key features of DeCompress:
  - Unsupervised: does not require clean-noisy image pairs for training
  - Data-efficient: Can operate with access to limited training data (only a single noisy image)
  - High-performance: Achieves state-of-the-art performance among zero-shot denoisers (for both Gaussian and Poisson noise)

DeCompress: Training phase

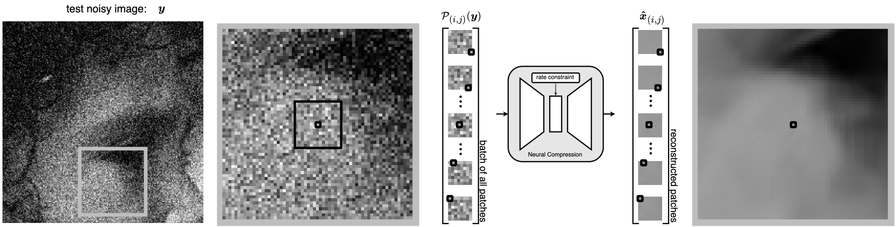
Step 1. Overlapping patches (e.g. 8×8) are extracted from a noisy image.

Step 2. A neural compression network is trained on extracted noisy patches.



DeCompress: Denoising phase

1. Extract overlapping patches from test image (`torch.nn.Unfold`).
2. Compress the batch of patches using the trained neural compressor (in parallel).
3. Combine and average all patches to reconstruct the denoised image (`torch.nn.Fold`).



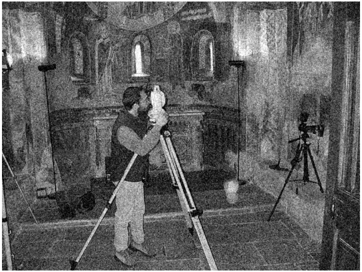
DeCompress variants

Two neural compression networks are trained for denoising without access to the ground-truth.

DeCompress (BSD400)

A set of 400 grayscale images of size 180×180.

DeCompress (single)



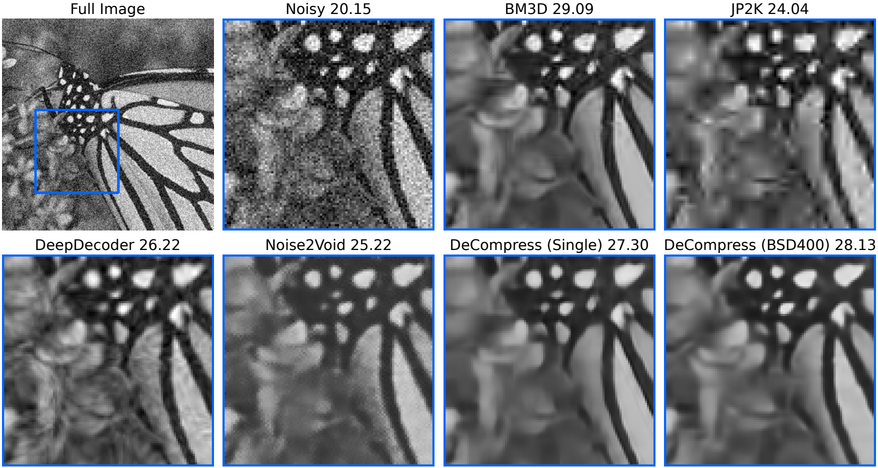
An image from *Imagenet* validation set. (`ILSVRC2012_val_00000059.png`)

Denoising peroformance comparison

Denoising performance on *Set11* images.

	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
noisy image	24.61	20.18	14.16
BM3D	**32.20**	**29.76**	**26.51**
JPEG2K	27.39	24.91	22.07
Deep Decoder	27.49	25.74	23.34
Noise2Void	26.88	24.77	19.88
DeCompress (single)	29.67	27.24	24.41
DeCompress (BSD400)	30.15	27.64	24.90

DeCompress: Sample visual comparison



Key takeaways & future directions

- Neural-compression-based denoising provides an effective denoising solution, especially in applications with little or no training data.
- Opens a new research direction: Exploring application of neural compression methods in solving extensive inverse problems.

Zero-shot Denoising via Neural Compression: arXiv:2506.12693

References

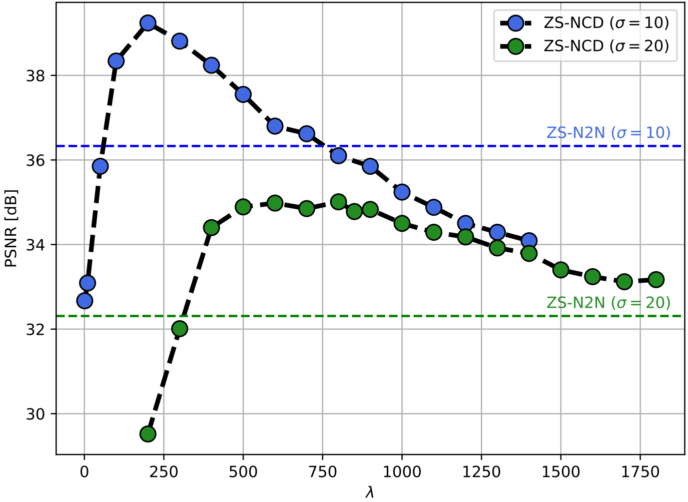
Donoho, David Leigh [2002]. *The kolmogorov sampler*. Department of Statistics, Stanford University.

Weissman, Tsachy and Ordentlich, Erik [2005]. "The empirical distribution of rate-constrained source codes". In: *IEEE transactions on information theory*.

Backup slides

A set of extra slides on zero-shot denoising via neural compression.

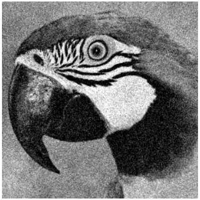
Effect of λ on denoising performance



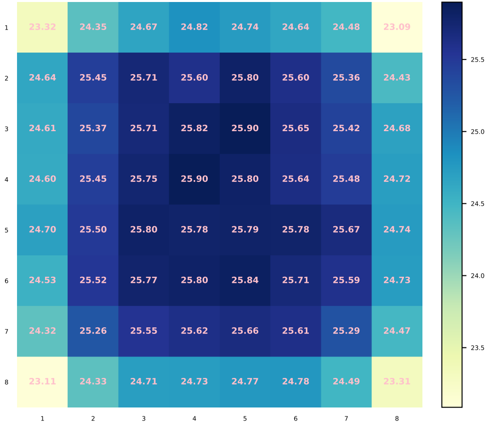
Effect of averaging

- Patch size: 8×8.
- PSNR achieved with only keeping single compressed pixel (no averaging).

Noisy Parrot (20.15 dB)



DeCompress (28.15 dB)



Zero-shot Denoising via Neural Compression

Let $\mathcal{Q} \subset \mathbb{R}^n$ denote the signal class of interest, such as vectorized natural images of a fixed size. A lossy compression code for $\mathcal{Q}$ is defined by an encoder-decoder pair (f, g), $f : \mathcal{Q} \to \{1, \ldots, 2^R\}$, and $g : \{1, \ldots, 2^R\} \to \mathbb{R}^n$ characterized by: i) *Rate R*, and ii) *Distortion δ*:

$$\delta = \sup_{\boldsymbol{x} \in \mathcal{Q}} \frac{1}{n} \|\boldsymbol{x} - g(f(\boldsymbol{x}))\|_2^2.$$

The set of reconstructions forms the codebook:

$$\mathcal{C} = \{g(i) : i = 1, \ldots, 2^R\} \subset \mathbb{R}^n.$$

Zero-shot Denoising via Neural Compression - Cont'd

We propose compression-based denoising as a structured maximum likelihood (ML) estimation. Given $\boldsymbol{y}$ and a lossy compression code (f, g) for $\mathcal{Q}$, the compression-based ML denoiser solves

$$\hat{\boldsymbol{x}} = \arg\min_{\mathbf{c}\in\mathcal{C}} \mathcal{L}(\mathbf{c}; \boldsymbol{y}),$$

where $\mathcal{L}(\mathbf{c}; \boldsymbol{y}) := -\sum_{i=1}^{n} \log p(y_i \mid c_i)$.

- Clean signals are more compressible than their noisy counterparts, by virtue of their structure.
- The most likely codeword from a clean-signal codebook naturally performs denoising.

Zero-shot Denoising via Neural Compression - Cont'd

1. AWGN. $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 I_n)$,

   $$\hat{\boldsymbol{x}} = \arg\min_{\mathbf{c} \in \mathcal{C}} \|\boldsymbol{y} - \mathbf{c}\|_2^2. \quad (1)$$

   Denoising by projecting the noisy observation onto the nearest codeword.

2. Poisson noise. $y_i \sim \mathrm{Poisson}(\alpha x_i)$ (low-light and photon-limited imaging scenarios)

   $$\hat{\boldsymbol{x}} = \arg\min_{\mathbf{c} \in \mathcal{C}} \sum_{i=1}^{n} \left(\alpha c_i - y_i \log c_i\right). \quad (2)$$

Theoretical findings

Theorem (AWGN)

Assume that $\boldsymbol{x} \in \mathcal{Q}$ *and let* (f, g) *denote a lossy compression for* $\mathcal{Q}$ *that operates at rate* R *and distortion* δ. *Consider* $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$, *where* $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 I_n)$. *Let* $\hat{\boldsymbol{x}}$ *denote the output of the compression-based denoiser defined by* (f, g) *as in* (1). *Then,*

$$\frac{1}{\sqrt{n}} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq \sqrt{\delta} + 2\sigma_z \sqrt{\frac{(2 \ln 2) R}{n}} (1 + 2\sqrt{\eta}),$$

with a probability larger than $1 - 2^{-\eta R + 2}$.

Corollary (Sparse Signals)

Let $\mathcal{Q}_n$ *denote the set of* k*-sparse vectors in* $\mathbb{R}^n$ *satisfying* $\|\boldsymbol{x}\|_2 \leq 1$. *Fix a parameter* $\eta \in (0,1)$, *and suppose* $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$ *where* $\boldsymbol{z} \sim \mathcal{N}(0, \sigma_z^2 I_n)$. *Then, there exists a family of compression codes such that, when used with the denoiser defined in* (1), *the estimate* $\hat{\boldsymbol{x}}$ *satisfies*

$$\frac{1}{n\sigma_z^2}\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2 \leq C \cdot \frac{k \log_2 n}{n} + \gamma_n,$$

with probability at least $1 - \frac{4}{(kn^{3k/2})^\eta}$. *Here,* $\gamma_n = o(1)$ *and* $C = 4 \ln 2(1 + 2\sqrt{\eta})^2$.

Theoretical findings - Cont'd

Theorem (Poisson)

Assume that $\mathbf{x} \in \mathcal{Q}$ *and let* (f, g) *denote a lossy compression for* $\mathcal{Q}$ *that operates at rate* R *and distortion* δ. *Assume that for any* $\mathbf{x} \in \mathcal{Q}$, $x_i \in (x_{\min}, x_{\max})$, *where* $0 < x_{\min} < x_{\max} < 1$. *Assume that* $y_1, \ldots, y_n$ *are independent with* $y_i \sim \mathrm{Poisson}(\alpha x_i)$. *Let* $\hat{\mathbf{x}}$ *denote the solution of* (2). *Let* $C_1 = x_{\max}^5/(x_{\min}^2)$ *and* $C_2 = \frac{x_{\max}^2}{x_{\min}^3}\beta\sqrt{(\frac{4}{\ln 2})}(\sqrt{1+\eta}+\sqrt{\eta})$. *Then, with a probability larger than* $1 - 2^{-\eta R + 2}$,

$$\frac{1}{n}\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq C_1\delta + C_2\sqrt{\frac{R}{n\alpha}}. \tag{3}$$