Evaluating Unsupervised Denoising Requires Unsupervised Metrics

NEW YORK UNIVERSITY

Adrià Marcos Morales[1,2], Matan Leibovich[3], Sreyas Mohan[1], Joshua Lawrence Vincent[4], Piyush Haluai[4], Mai Tan[4], Peter Crozier[4], Carlos Fernandez-Granda[1,3]

[1]Center for Data Science, New York University, New York, NY [2] Centre de Formació Interdisciplinària Superior, Universitat Politècnica de Catalunya, Barcelona, Spain [3] Courant Institute of Mathematical Sciences, New York University, New York, NY [4] School for Engineering of Matter, Transport & Energy, Arizona State University, Tempe, AZ

Motivation and Overview

- **Unsupervised denoising** is a crucial challenge in real-world imaging applications
- Deep-learning methods have demonstrated **impressive performance** on benchmarks based on **synthetic noise**
- **No metrics exist** to evaluate these methods *in an unsupervised fashion*
- This is problematic because in many practical applications **ground-truth clean images** are simply **not available**
- We propose **two novel metrics**: the unsupervised mean squared error (**uMSE**) and the unsupervised peak signal-to-noise ratio (**uPSNR**), which are **computed using only noisy data**.
- These metrics are proven to be **asymptotically consistent estimators** of supervised counterparts, MSE and PSNR.
- We evaluate uMSE and uPSNR via controlled numerical experiments with synthetic noise and real-world electron-microscopy data

Proposed Metrics

We compare the denoised image to a noisy reference *with the same underlying clean signal*

A **correction term** (based on two additional noisy references) neutralizes the noise contribution

Unsupervised mean squared error (uMSE): Given a noisy input signal $y \in \mathbb{R}^n$ and three noisy references $a, b, c \in \mathbb{R}^n$ the unsupervised mean squared error of a denoiser $f : \mathbb{R}^n \to \mathbb{R}^n$ is

$$\mathrm{uMSE} := \frac{1}{n}\sum_{i=1}^{n}(a_i - f(y)_i)^2 - \frac{(b_i - c_i)^2}{2} \qquad (1)$$

Unsupervised peak signal-to-noise ratio (uPSNR): Given an uMSE value computed from y, a, b, $c \in \mathbb{R}^n$ the unsupervised peak signal-to-noise ratio of a denoiser $f : \mathbb{R}^n \to \mathbb{R}^n$ is

$$\mathrm{uPSNR} := 10 \log\left(\frac{\mathrm{M}^2}{\mathrm{uMSE}}\right), \qquad (2)$$

where M is the maximum possible value of the signal of interest

Theoretical Guarantees

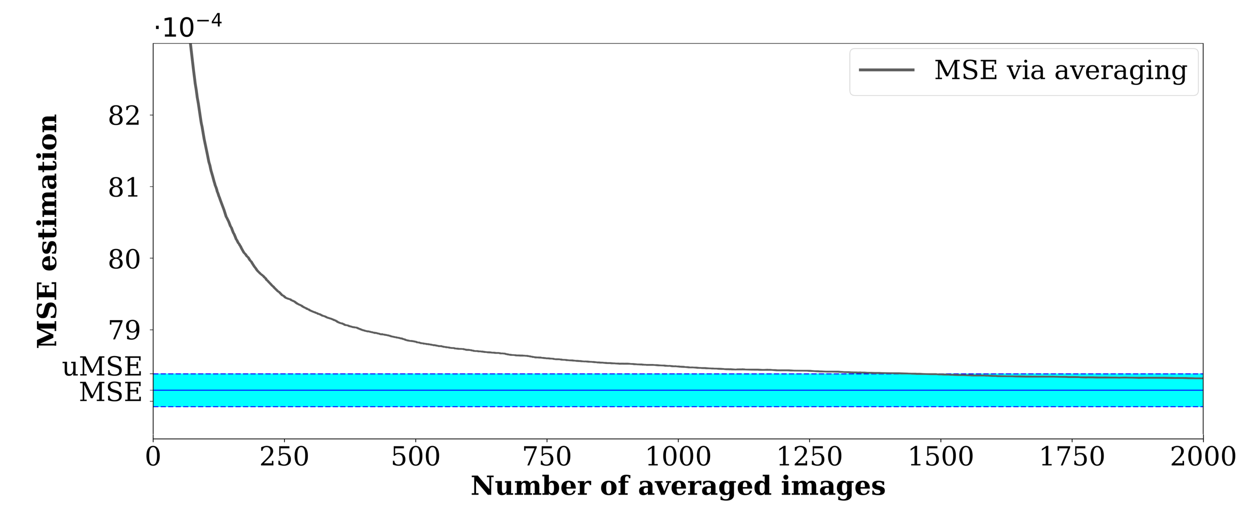
If the noisy references correspond to the *same clean image* and the noise is *pixel-wise independent*

- The uMSE and uPSNR are **unbiased** (their means equal the MSE and PSNR)
- The uMSE and uPSNR are **consistent** (they converge to the MSE and PSNR)
- The uMSE is asymptotically **Gaussian**, which enables the construction of confidence intervals

Comparison to Existing Approaches

In existing work, unsupervised methods are evaluated:

- On simulated data with known clean images (*Problem:* Not real noise)
- By visual inspection (*Problem:* Not quantitative)
- referencesring to *clean* images estimated via averaging (*Problem:* Many noisy copies are needed)



Computing Noisy Realizations in Practice

Challenge: How to compute noisy references with *same clean image* and *independent noise*?

- **Multiple images:** References correspond to *consecutive frames* acquired within a short time interval (inspired by Noise2Noise). Preferable when image content does not experience rapid dynamic changes from frame to frame
- **Single image:** References computed from a single image via spatial subsampling. Pixels in each 2×2 block are assigned to each reference (inspired by Neighbor2Neighbor)

Noisy image | Subsampled references | Subsampling scheme

WARNING: The references should be constructed to ensure that (1) the underlying clean signal is as similar as possible, (2) the noise is independent across references. Violating these assumption may **bias** the estimators

[1] J. Batson and L. Royer. **Noise2self: Blind denoising by self-supervision.** 2019
[2] T. Huang et al. **Neighbor2neighbor: Self-supervised denoising from single images.** 2021
[3] J. Lehtinen et al. **Noise2noise: Learning image restoration without clean data.** 2018
[4] A. Krull et al. **Noise2void - learning denoising from single noisy images.** 2019.
[5] D. Y. Sheth, S. Mohan et al. **Unsupervised deep video denoising.** 2021
[6] S. Laine et al. **High-quality self-supervised deep image denoising.** 2019

Results

Controlled experiments show that uMSE and uPSNR provide accurate approximations to the MSE and PSNR, without access to clean ground-truth images

	Natural images with Gaussian noise ($\sigma = 50$)			Electron-Microscopy images with Poisson Noise		
Method	PSNR	uPSNR	uPSNR$_S$	PSNR	uPSNR	uPSNR$_S$
Bilateral	21.84	21.86	22.90	20.18	20.20	20.21
UNet	24.95	24.96	23.52	24.65	24.69	24.79
DnCNN	23.95	24.0	26.08	25.74	25.68	25.86
BlindSpot	24.08	24.07	22.77	24.86	24.87	24.74

Single-image uPSNR (uPSNR$_S$) is more accurate for electron-microscopy images because they are smoother (so the spatially subsampled clear images are more similar)

Real-World Application: Denoising at the Atomic Level

Unsupervised denoising is crucial for **transmission electron microscopy** (TEM), a key imaging modality in material sciences. We apply unsupervised denoising to 18,597 noisy TEM frames depicting **platinum nanoparticles on a cerium oxide support** and use uPSNR to evaluate them