Mitigating Topological Noise in 3D Images of Porous Media

NJIT Grace Hopper AI Research Institute Seed Program

Aakash Karlekar, Ebru Dagdelen, Manav Arora, Matthew Illingworth, Jonathan Jaquette, Linda J. Cummings, Lou Kondic

DMS-2201627
DMR-2410985

Introduction

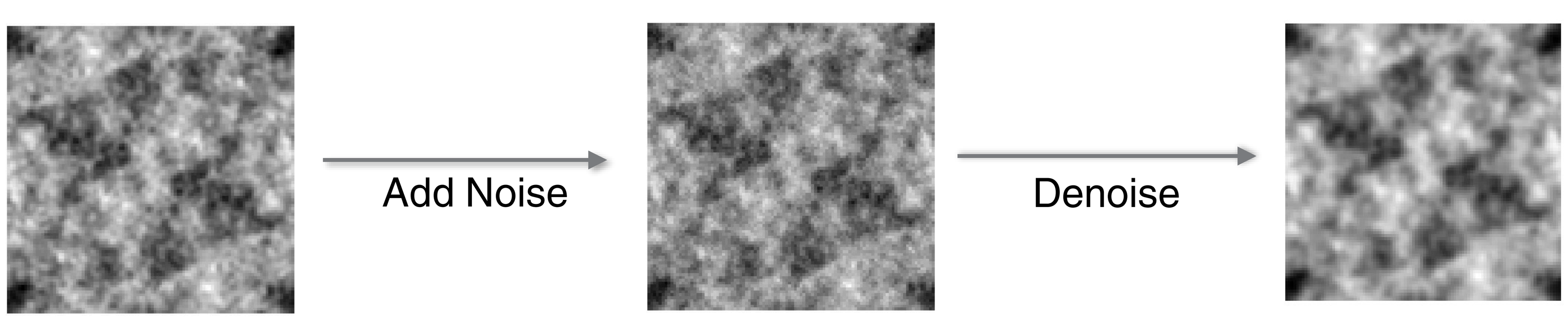
Flow through porous media is essential to many natural and industrial processes, including environmental cleanup, oil recovery, and CO_2 storage. Understanding and optimizing such processes requires characterizing the complex internal pore structure of these materials. Techniques from **Topological Data Analysis** (TDA), especially persistent homology, are very helpful for this task.

However, when working with real-world data — specifically 3D images of porous media — we face significant computational challenges because of the complexity of the image data and the presence of experimental noise, which can obscure key topological features and increase computational costs. To address these challenges, it becomes essential to develop methods that reduce computational complexity and mitigate noise present in the image data, while still preserving its key features.

Effective resolution of this challenge is an active area of ongoing research. Methods ranging from Gaussian filters to deep learning approaches have been explored as potential ways to denoise noisy image data prior to analyzing it [1]. We employ this approach, using a smoothing method to denoise noisy porous media image data, and primarily focus on formulating several measures that allow us to verify that we have strictly removed noise from our data rather than essential information.

Methodology

For our smoothing method, we employ a standard denoising method using Gaussian convolution with a smoothing parameter σ to smooth the data and reduce noise. We devise a testing process where we start out with simulated image data, which we refer to as the original dataset. We then add controlled static noise from a known distribution to this dataset to produce a noisy dataset. Next, we apply our smoothing method on the noisy dataset for various values of the smoothing parameter σ, referring to the resulting dataset after smoothing as the denoised dataset.



Original Dataset **Noisy Dataset** **Denoised Dataset**

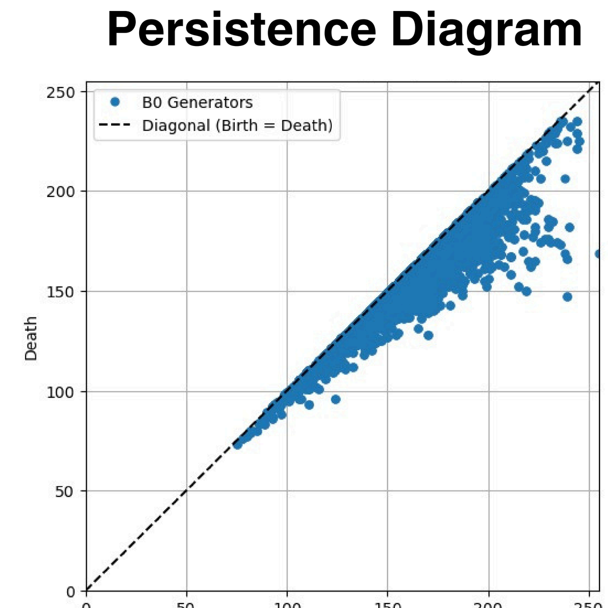
We identify several normalized topological measures that we use to quantify the difference between the original and denoised datasets. These measures include absolute differences in some persistence statistics computed for the original and denoised datasets, the bottleneck and Wasserstein distances between **Persistence Diagrams** (PD) of the original and denoised datasets, and the L_2 norm distances between various persistence representations of the original and denoised datasets. We use these measures to evaluate the efficacy of our denoising method.

Topological Measures and Representations

Persistent Homology (PH) is a key method in TDA that can be used to quantify a medium's heterogeneity across various scales. In our work, we start with 3D grayscale images of porous media, where each voxel takes on an integer value on the interval [0, 255] assumed to correlate with the permeability of the material being studied, with 0 corresponding to void space and 255 corresponding to solid material. Then, through a thresholding process based on the grayscale level of each voxel in the image, we construct a nested sequence of geometric layers called a filtration. We can study the topological properties of each of these filtration layers by computing its relevant homology groups, which are simply described through the Betti numbers – β_0, β_1, β_2 – which correspond to the number of connected components, loops, and cavities found within a filtration layer.

PH looks to synthesize this information by quantifying the topological features that persist across many layers of a filtration. These features are referred to as generators. The threshold value at which a generator appears is referred to as its birth number, while the threshold value at which a generator disappears is referred to as the generator's death number. The difference between the birth and death number is referred to as the lifespan of the generator. We track persistence statistics such as the number of generators encountered within a filtration and the average lifespan of these generators.

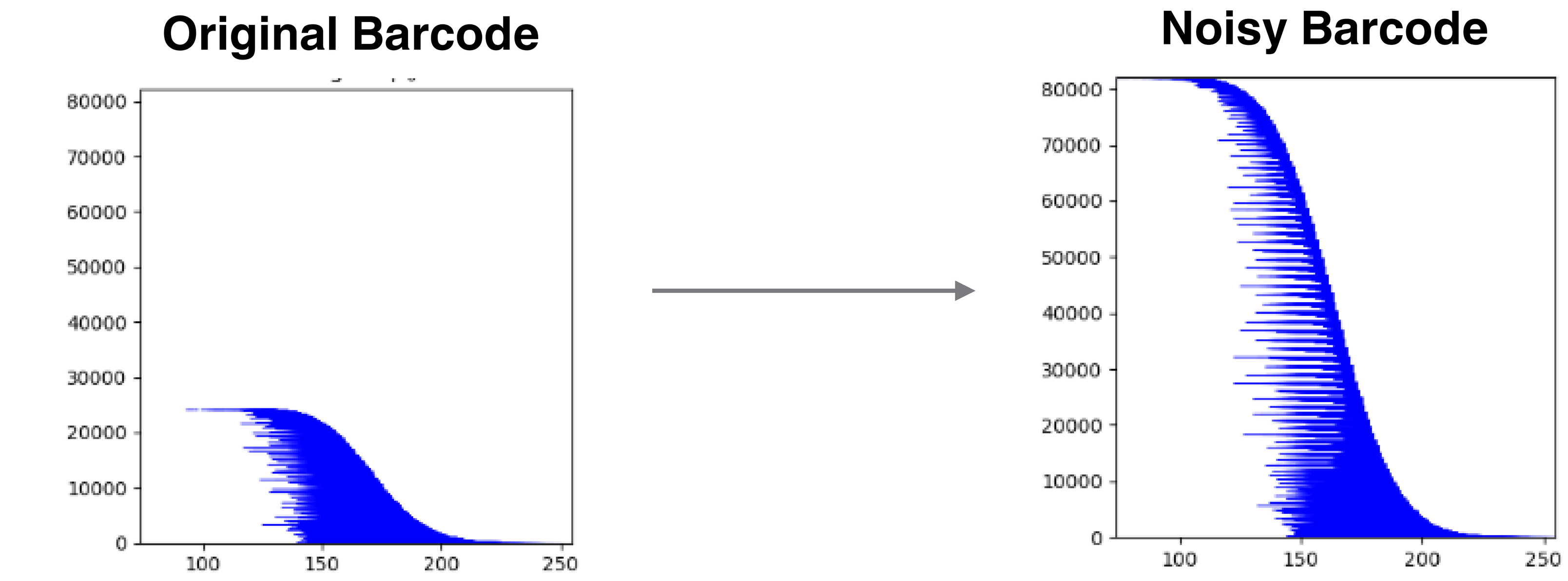
Persistence Diagram



We can visualize the persistence of each generator as a point in a 2D graph, where one axis represents a generator's birth number and the other its death number. By plotting the corresponding birth-death pairs for the generators we find in our filtration, we obtain a PD.

We can use a PD to generate other persistence representations. One such representation, **Persistence Barcodes** allow us to visualize the lifespan of a porous material's topological features using bars whose lengths correspond to the lifespan of the feature. When noise is introduced to a porous media image, there is a sharp increase in the number of short-lived topological features. This impact is evidenced through the larger spread of the persistence barcodes for the noisy dataset as compared to the persistence barcodes for the original dataset.

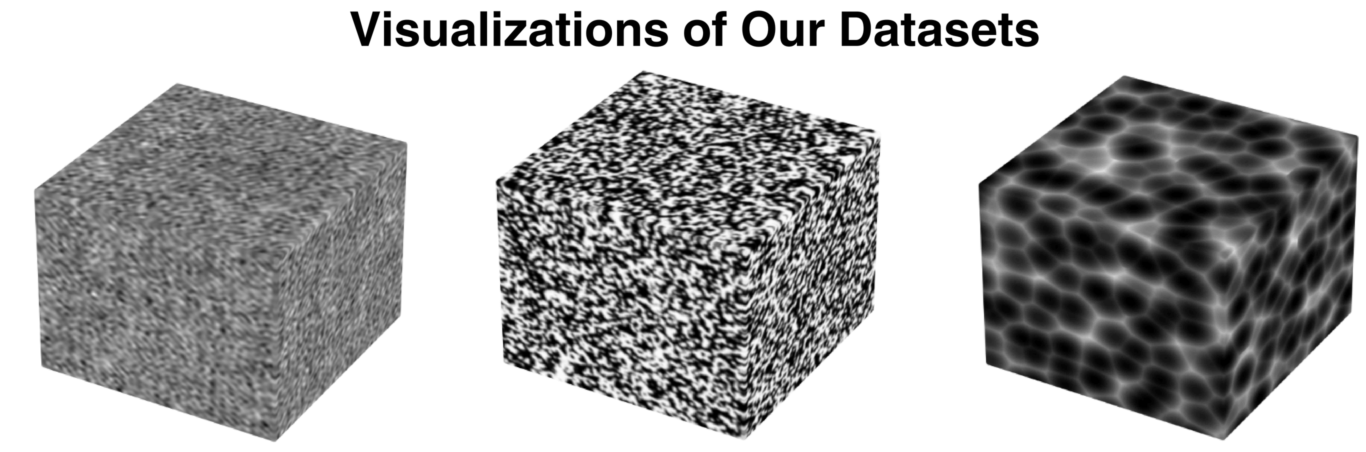
Original Barcode **Noisy Barcode**



Results

To analyze our denoising method, we apply our testing process to several datasets: A synthetic 3D image generated using Fourier series, which we refer to as the **Fourier dataset;** a dataset consisting of random, overlapping spheres generated using the PuMA library which we refer to as the **PuMA dataset** [2]; and a Worly noise based dataset which we refer to as the **cellular dataset** [3].

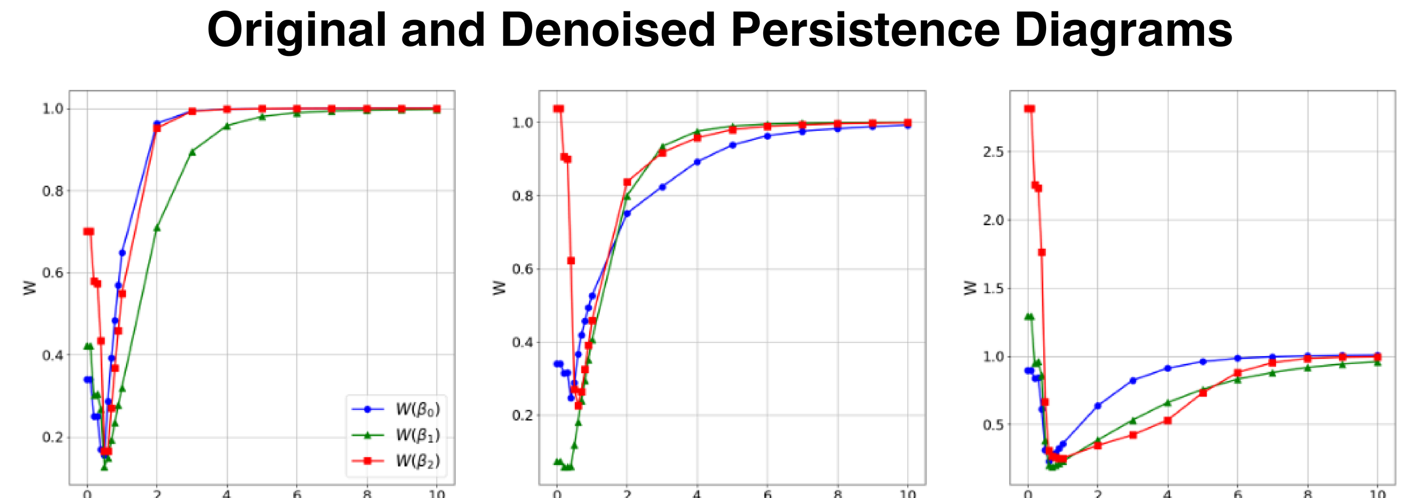
Visualizations of Our Datasets



Fourier **PuMA** **Cellular**

We then analyze the efficacy of our denoising method through the lens of several normalized measures based on persistence statistics, such as the number of generators and average lifespan of generators, as well as various persistence representations. From these measures, we are able to ascertain that our method is able to minimize the overall difference between the denoised and original versions of each of our datasets for some ideal denoising parameter, consistently found to be around σ ≈ 0.5.

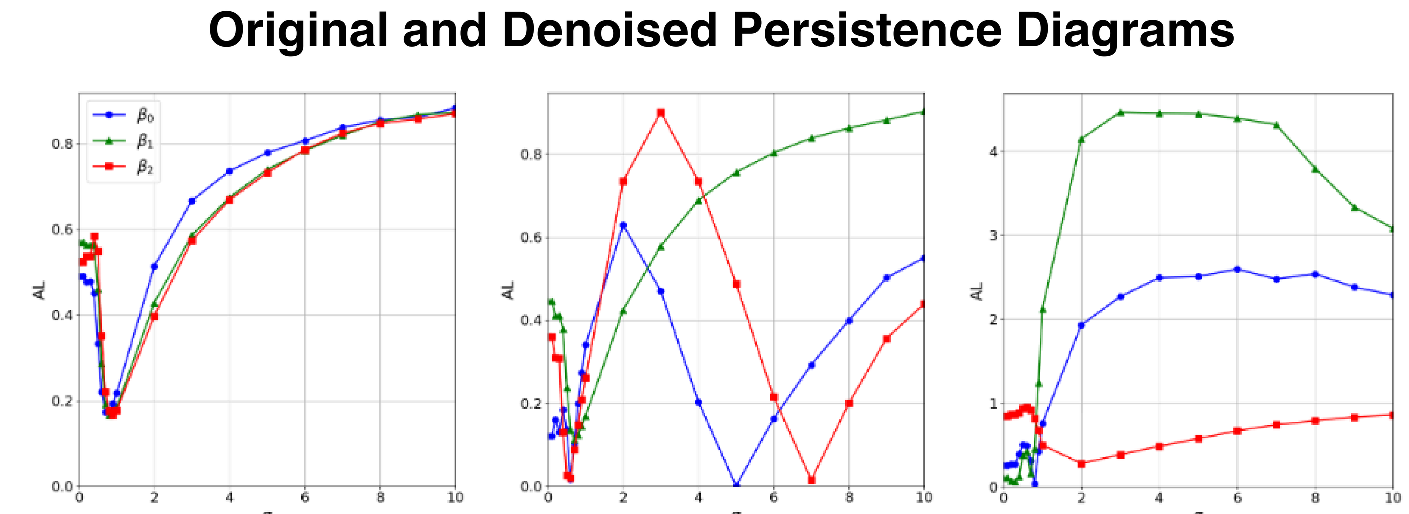
Normalized Wasserstein Distance (W) between Original and Denoised Persistence Diagrams



Fourier **PuMA** **Cellular**

We also find that some of our measures are more robust than others. For instance, the bottleneck distance measure is less susceptible to noise as compared to the Wasserstein distance measure and the average lifespan measure is not as stable/smooth as compared to our other measures.

Normalized Average Lifespan Measure (AL) between Original and Denoised Persistence Diagrams



Fourier **PuMA** **Cellular**

References

[1] R. Turke˘s, J. Nys, T. Verdonck, S. Latr´e, Noise robustness of persistent homology on greyscale images, across filtrations and signatures, PLOS ONE 16 (9) (2021) 1–26.
[2] J. C. Ferguson, F. Semeraro, J. M. Thornton, F. Panerai, A. Borner, N. N. Mansour, Update 3.0 to "puma: The porous microstructure analysis software",(pii: S2352711018300281), SoftwareX 15 (2021) 100775.
[3] R. A. McLeod, pyfastnoisesimd, https://github.com/robbmcleod/pyfastnoisesimd, accessed: 2025- 08-04.