RANDOM FILTERS FOR ENRICHING THE DISCRIMINATORY POWER OF TOPOLOGICAL REPRESENTATIONS

Grayson Jorgenson, Henry Kvinge [*], and Tegan Emerson [†]
Mathematics of Data Science Team
Pacific Northwest National Laboratory
Seattle, Washington
{grayson.jorgenson,henry.kvinge,tegan.emerson}@pnnl.gov

Colin Olson
U.S. Naval Research Laboratory
Washington, D.C.
colin.olson@nrl.navy.mil

ABSTRACT

Topological representations of data are inherently coarse summaries which endows them with certain desirable properties like stability but also potentially inhibits their discriminatory power relative to fine-scale learned features. In this work we present a novel framework for enriching the discriminatory power of topological representations based on random filters and capturing "interference topology" rather than direct topology. We show that our random filters outperform previously explored structured image filters while requiring orders of magnitude less computational time. The approach is demonstrated on the MNIST dataset but is broadly applicable across data sets and modalities. This work is concluded with a discussion of the mathematical intuition underlying the approach and identification of future directions to enable deeper understanding and theoretical results.

1 INTRODUCTION

One of the fundamental challenges of machine learning is developing methods to distill a potentially complex and high-dimensional dataset into a set of features that can be used to perform a specific task. Deep learning models have been successful in a range of problem spaces, from image recognition to object detection LeCun et al. (2015), due to their ability to learn high-performance, analysis-dependent features. In the last decade there has been an explosion of work in topological data analysis Carlsson (2009), a relatively new field that explores the application of topology-inspired statistics to problems in data science. Unlike the features extracted by deep learning algorithms, the mathematical field of topology studies coarse properties associated to shape. Topological features have a range of attractive properties and there has been recent interest (for example, Kindelan et al. (2021)) in explicitly merging the benefits of topologically-inspired features with the discriminatory power of deep learning frameworks. The present work is a step toward building this connection.

In this work we narrow our focus to the problem of extracting topological features from image data. We employ the *lower star filtration* (LSF), a common method for extracting topological features from grayscale images, that creates simplicial complexes based on relative pixel intensities. One contribution of this work is a method for addressing challenges that arise when working with image data expressing low topological variation. Specifically, we address the challenge when the LSF on its own fails to extract enough features to distinguish between different classes in a supervised learning task. A concrete example of this is the MNIST dataset LeCun et al. (1998) where, for example, the digits "0","6", and "9" are topologically equivalent. We introduce a novel approach that blends a

[*]Dr. Kvinge holds a joint appointment in the Department of Mathematics at the University of Washington.
[†]Dr. Emerson holds joint appointments in the Department of Mathematics at Colorado State University and the Department of Mathematical Sciences at the University of Texas, El Paso.

given image with a collection of random images (formed by applying a Gaussian blurring kernel to a uniform random noise matrix) to enrich the discriminatory power of topological features captured from the lower star filtration.

Intuitively, rather than capturing the topological structure demonstrated in the base image, we look at how the base image alters topological structures across the set of secondary images. We demonstrate that the pipeline comprised of *random filters* and LSF, summarized by a set of persistence images (PIs) Adams et al. (2017), is able to improve classification performance on MNIST. Unlike the approach presented in Garin & Tauzin (2019) that leverages computationally expensive, structured filters and random forest classifiers, this work integrates the topological representation into a convolutional neural network (CNN) classification model applied to a multi-channel PI formed from multiple filters. Further, we show that topological features provide a stronger starting point for the CNN-based classification rather than the raw grayscale MNIST images.

The remainder of this paper is structured as follows. First, in Section 2 we present a brief primer on topological data analysis including an introduction to the LSF. Next, in Section 3 we present our novel, random filter approach. Section 4 presents computational results for the MNIST dataset. Finally, we conclude with a discussion of our findings and identification of next steps for formalizing and building theoretical understanding of the developed approach.

2 A BRIEF PRIMER ON TOPOLOGICAL DATA ANALYSIS

A fundamental tool in the application of topology to data science is *persistent homology*, first introduced in Edelsbrunner et al. (2000). This tool tracks the formation and vanishing of features along a given nested sequence of topological spaces, capturing the features as generators in the corresponding persistent homology groups. The set of pairs of indices in the nested sequence where features first appear and disappear (colloquially referred to as *birth* and *death* times) is known as a *persistence diagram* (PD) and provides a convenient way of recording the feature information by simply tracking how long they “persist” in the sequence. In the case of grayscale image data, it is common to consider the image as a discrete sampling of 2-dimensional function and define the topology based on the levelsets of that function using the LSF. Figures 1 and 2 provide visualizations of the LSFs[1] for unfiltered and filtered MNIST images, respectively. The corresponding 0- and 1-dimensional PDs for each LSF are also provided. These PDs are then converted to PIs as outlined in Adams et al. (2017)[2].

3 RANDOM FILTERS FOR ENRICHING TOPOLOGICAL REPRESENTATIONS

Our technical approach addresses the setting when there is insufficient topological variation to achieve strong discrimination across a data set by extracting topological features from the raw data alone. Particularly, we focus on the MNIST dataset where the LSF computed directly from the grayscale images fails to produce sufficiently discriminatory topological representations. In Garin & Tauzin (2019) the authors were able to successfully classify MNIST images with a random forest using features extracted from a similar LSF by diversifying the topological structure through the application of structured filters; they did this by first applying several traditional binary-to-grayscale image filters to binarizations of the MNIST images. However, some of those filters can be expensive to compute and have limited diversity. With this in mind, we introduce the following family of computationally inexpensive filters.

Building on the foundation laid in Garin & Tauzin (2019), we create a more flexible framework building on random rather than structured filters. We define a *random filter* to be an $m \times n$ grayscale image formed by convolving a random matrix (with entries sampled uniformly from $[0, 1]$) with a Gaussian blurring kernel. A random filter, R, is applied to an arbitrary $m \times n$ grayscale image, I, from the dataset via elementwise multiplication; the filtered image is $R \odot I$. In the specific situation

[1]All results shown herein were produced using the LSF implementation from the `Ripser` Tralie et al. (2018) persistent homology Python package.

[2]All persistence images we use here are calculated with the `Persim` package that is part of `Scikit-TDA` Saul & Tralie (2019). Given that we are considering grayscale images, the maximum birth and persistence scales are naturally bounded to [0,1]. We choose a Gaussian variance of $\sigma^2 = 0.003$.

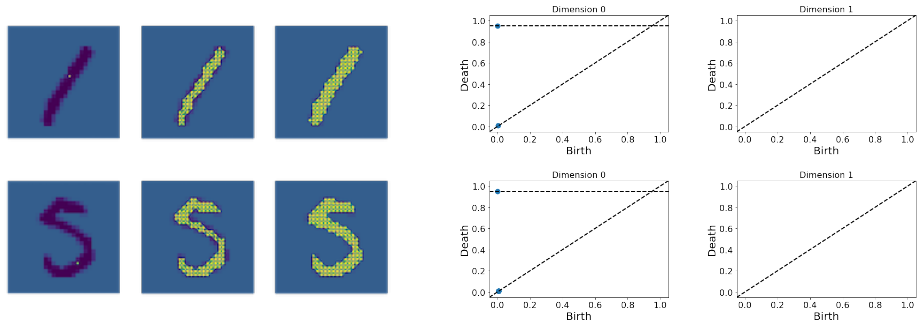

Figure 1: Complexes in the LSF of the raw, inverted MNIST images and their persistence diagrams. There are three complexes for each digit and from left to right the threshold values are $0.0, 0.4$, and 0.8, respectively. The corresponding PDs for dimensions $0, 1$ are on the right.

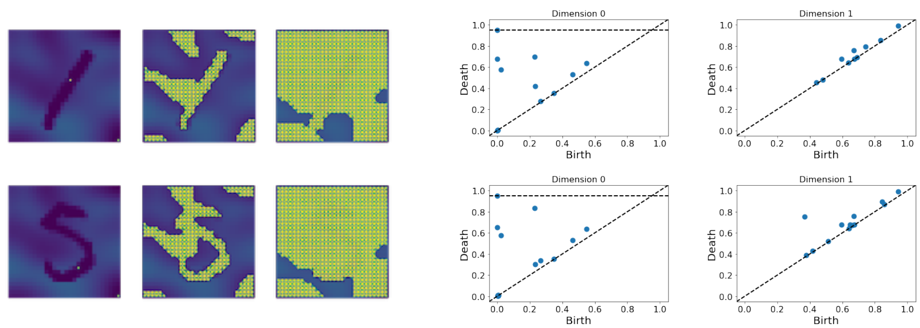

Figure 2: Complexes in the LSF of the filtered, inverted MNIST images and their persistence diagrams. There are three complexes for each digit and from left to right the threshold values are $0.0, 0.4$, and 0.8, respectively. The corresponding PDs for dimensions $0, 1$ are on the right.

of MNIST, the pixel variation induced by a random filter helps the shape of each digit leave an "imprint" on the topological information of that filter. Figure 2 shows this in action, repeating the experiment of Figure 1 but for the MNIST images after application of a random filter (Figure 3).

It can be seen in Figure 3 that the shape of the digit 5 enables a loop to form early on while no loops form in the corresponding complex for the digit 1 at that time; this is a distinction between the digits that is now captured by the topological representation. We further boost the discriminatory power by forming N random filters, $R_1, \dots, R_N$. For each grayscale image I in the dataset, we create the dimension $0, 1$ PIs of each $R_i \odot I$. Thus, I is represented by $2N$ PIs total and these are stacked to create the input tensor (that is, a multi-channel PI) for the CNN. The formation of a multi-channel PI for integration with CNN classifiers is a further novel contribution of this work.

While this work only considers application of random filters to grayscale images the intuition is widely applicable: in the absence of sufficient topological variation across classes one may instead consider the notion of "interference topology" to achieve stronger discrimination across classes.

4 Classifying MNIST with random filters and persistence images

We examine the effectiveness of our preprocessing approach on the MNIST dataset consisting of $60,000$ training samples of handwritten digits and $10,000$ test samples. Since MNIST is a well-solved classification task with small CNNs like LeNet able to get $\approx 99\%$ test accuracy, we compare the effectiveness of the same CNN architecture applied to both the raw MNIST images (of size 28×28 pixels) and our multi-channel PI for the purpose of benchmarking our TDA pipeline. To that end, the results presented in Table 1 were produced using a simple CNN architecture consisting

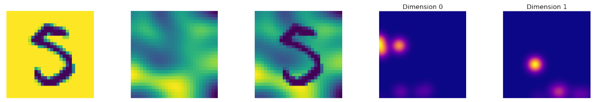

Figure 3: Visualizations of persistence images for an MNIST sample with a random filter applied. Left to right: the raw inverted MNIST image, the smoothed uniform noise image, the filtered MNIST image, the dimension 0 PI, and the dimension 1 PI.

Table 1: Accuracies on MNIST. Each model used the architecture described in Section 4 and was trained for six epochs. The reported accuracies are the means and standard deviations over five random weight initializations. The highest accuracy in indicated in bold.

Input type	Test accuracy (%)
Raw MNIST image input	91.16 ± 0.28
25 Structured filters in Garin & Tauzin (2019), 7×7 resolution PIs	95.33 ± 0.11
25 random filters, 7×7 resolution PIs (ours)	97.17 ± 0.1
80 random filters, 7×7 resolution PIs (ours)	$\mathbf{98.2 \pm 0.05}$

of a single convolutional layer with 1×1 convolutional filters, a ReLU activation layer, and a single fully connected layer. No pooling or dropout is used.

It was shown in Adams et al. (2017) that PIs are robust to the choice of resolution. We explored the trade-off between PI resolution and the number of random filters on classification accuracy. The results of this experiment are provided in Subsection 4.1 and led us to choose to use 7×7 resolution PIs. We report the accuracies of three topological methods in Table 1: the accuracy from using PIs generated from the 25 structured filters (raw grayscale, inverted raw grayscale, height, radial, density, dilation, erosion, and signed distance) defined in Garin & Tauzin (2019) with their hyperparameter choices, and the accuracy from using PIs obtained from 25 and 80 random filters. Because we compute both the dimension 0 and dimension 1 persistence images for each filtered image, the total numbers of input channels into the CNN are 50 and 160, respectively. For reference, we also report the test accuracy from using the same architecture on the raw, 28×28 grayscale MNIST images.

From the results in Table 1 it is easily seen that for the same number of filters, the random filters achieve superior classification though both topological approaches outperform the same architecture applied to the raw data. Finally, we also see that there is increased performance gains for the larger number of random filters. In Table 2, we compare run times from generating the 25 structured image filters and 25 random filters. The structured filters are several orders of magnitude slower than our novel random filters. This is likely partly due to inefficiencies in our iterative implementations of the traditional filter algorithms. However, for the random filters the source smoothed noise images only need to be generated once and the application of each depends only on elementwise multiplication which is already highly optimized in Python libraries such as NumPy.

4.1 Persistence image resolution and number of random filters

In this subsection, we describe an experiment that identifies the effect of resolution and number of random filters on classification accuracy. We train CNNs of the same shape as the model used in Section 4 (that is, a single convolutional layer with 1×1 filters followed by a linear layer) on a small subset of MNIST consisting of $2,100$ evenly sampled training images and 900 evenly sampled testing images. For each of a range of possible persistence image resolutions, we train a model using peristence images of that resolution for different numbers of random filters. The mean accuracies over five trials are reported in Figure 4.

In all cases, using too few random filters (e.g., < 15) appears to be deleterious; more "perspectives" are needed to successfully distinguish between the MNIST images. Using a larger resolution also helps improve accuracy, though for both resolution and number of filters the benefit from increasing either is diminishing.

Table 2: Run time comparisons from generating 25 random filters and the 25 structured filters on subsets of MNIST of size 10, 100, and 1000 using an Ubuntu machine with 2.20 GHz processor speed.

	10 samples	100 samples	1000 samples
Structured filters	11.3s	1min 54s	19min 11s
Random filters	4.82ms	9.13ms	49.3ms

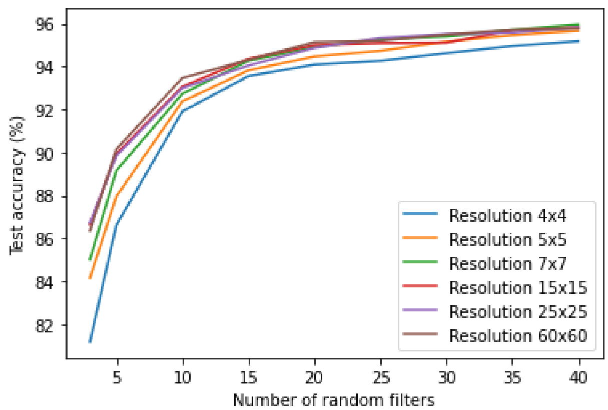

Figure 4: Plot of test accuracies on a subset of MNIST as the resolution of the PIs and the numbers of filters is changed.

We chose 7×7 resolution persistence images for the experiments in Section 4 as a compromise between accuracy and the size of the input tensor. For example, using 80 random filters and 30×30 persistence images would require using roughly 160 times the amount of RAM that MNIST images consume by themselves which can be impractical.

5 Discussion and Future Work

We have presented a novel approach for boosting the discriminatory power of topological features through the use of random filters. These random filters produce richer topological representations and higher MNIST classification accuracy than previously employed structured filters while dramatically reducing computational run times. This work is also contributes a novel approach for forming multi-channel PIs which can be integrated into CNN architectures. One avenue of future exploration will focus on the determination of optimal architectures for analysis of PIs.

This paper focused solely on the concrete example of the MNIST data set where the raw data has insufficient topological variation for discrimination. Our findings suggest that coarse topological features, when collected from multiple perspectives, provides richer, more discriminatory topological representations. Intuitively, this finding is likely related to random projections where one can understand data more comprehensively using many projections rather than one. Future efforts will formalize this connection to provide greater theoretical understanding and generalize the framework for widespread application.

References

Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18, 2017.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*, pp. 454–463. IEEE, 2000.

Adélie Garin and Guillaume Tauzin. A topological "reading" lesson: Classification of mnist using tda. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 1551–1556. IEEE, 2019.

Rolando Kindelan, José Frías, Mauricio Cerda, and Nancy Hitschfeld. Classification based on topological data analysis. *arXiv preprint arXiv:2102.03709*, 2021.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, pp. 436–444, 2015.

Nathaniel Saul and Chris Tralie. Scikit-tda: topological data analysis for python. *URL https://doi.org/10.5281/zenodo*, 2533369, 2019.

Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser. py: A lean persistent homology library for python. *Journal of Open Source Software*, 3(29):925, 2018.