Challenging Euclidean Topological Autoencoders

Michael Moor, Max Horn, Karsten Borgwardt, Bastian Rieck

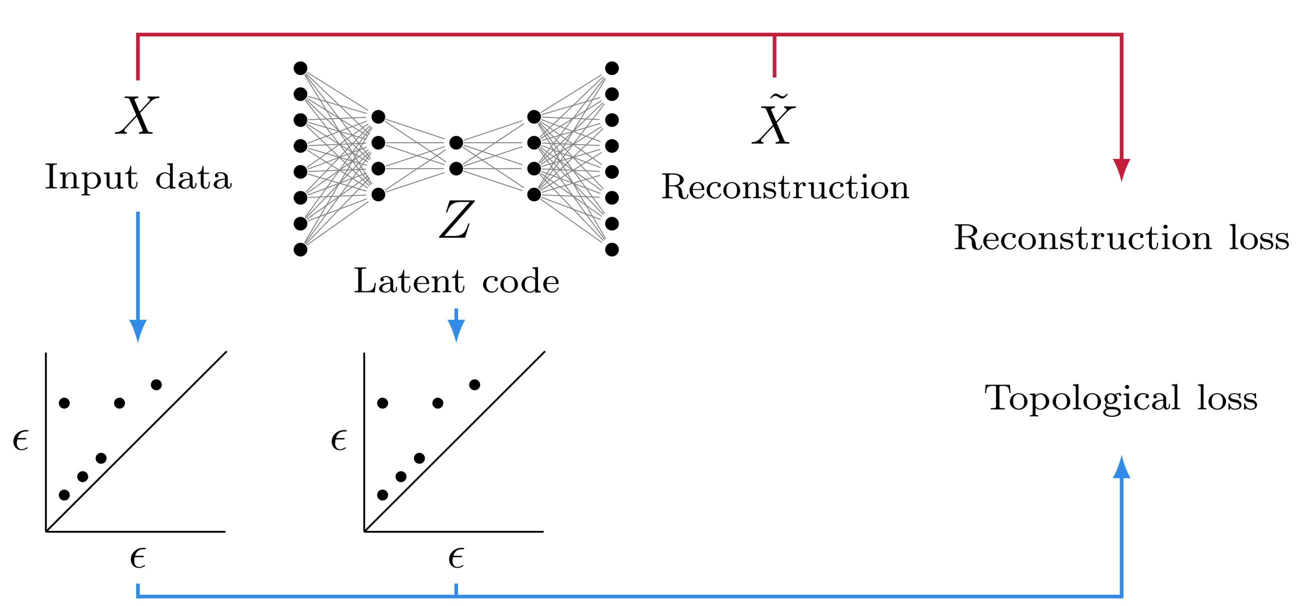
D-BSSE, Machine Learning and Computational Biology Lab, ETH Zurich, Switzerland

Summary

- Topological autoencoders (TopoAE) have demonstrated their capabilities for performing dimensionality reduction while at the same time preserving topological information of the input space.
- In its original formulation, this method relies on a Vietoris-Rips filtration of the data space, using the Euclidean metric as the base distance. It is commonly assumed that this distance is not sufficiently powerful to capture salient features of image data sets.
- Here, we investigate *alternative* choices of distances in the data space, which are generally considered to be more faithful for image data in comparison to the pixel distance.
- Our experiments on real-world image datasets show that the Euclidean formulation of TopoAE is surprisingly competitive with more elaborate, perceptually-inspired image distances as well as random convolutions.

Background: Topological autoencoders

Topological autoencoders (TopoAE) [1] have been recently proposed to equip autoencoders with a differentiable topological loss term which incentivises the latent encodings to preserve topological features of the data space.
Below, a mini-batch X is fed through a standard bottleneck architecture in order to recover the reconstruction $\tilde{X}$. Additionally, treating both the input X and the latent code Z as point clouds, a Vietoris-Rips filtration is computed based on the Euclidean metric. This results in two persistence diagrams which are provided to the topological loss term.



Topological Loss: $\mathcal{L}_t = \mathcal{L}_{\mathcal{X}\to\mathcal{Z}} + \mathcal{L}_{\mathcal{Z}\to\mathcal{X}}$

$$\mathcal{L}_{\mathcal{X}\to\mathcal{Z}} := \frac{1}{2}\left\|A^X\left[\pi^X\right] - A^Z\left[\pi^X\right]\right\|^2 \qquad \mathcal{L}_{\mathcal{Z}\to\mathcal{X}} := \frac{1}{2}\left\|A^Z\left[\pi^Z\right] - A^X\left[\pi^Z\right]\right\|^2$$

Measuring distance in the data space

For constructing the Vietoris-Rips complex, TopoAE employs the Euclidean distance $d(x_i, x_j) = \|x_i - x_j\|_2$. For images this amounts to a pixel distance which is typically assumed to be supoptimal in terms of faithfully capturing similarities between images (e.g. due to the lack of translation equivariance). Here, we empirically test this hypothesis by performing an ablation study including TopoAE and 2 alternative formulations that employ domain-specific image distances.

Random convolutions

Convolutional neural networks (CNNs) exhibit an inductive bias which is advantageous for the feature extraction and classification of images. Even randomly initialised convolutional layers can be used as rich image feature extractors. As we aim to preserve the input distance, learning the data distance simultaneously would be circular. Therefore, here we base the input distance on the feature map output $\mathcal{F}$ of an *untrained* CNN.

$$d(x_i, x_j) = \|\mathrm{vec}(\mathcal{F}(x_i)) - \mathrm{vec}(\mathcal{F}(x_j))\|_1$$

Perceptual similarity

Zhang et al. [2] showed that the feature representations of deep neural networks (DNNs) are well-suited to quantify the similarity of images similarly to human perception. We thus also consider a learnt similarity score based on DNN features called Learned Perceptual Image Patch Similarity (LPIPS) [2]. The LPIPS distance is derived by training a DNN, such that the difference of its features is in line with human perceptual scores.
Here, the distance between two images x_i and x_j is computed via their features $\hat{y}_i$ and $\hat{y}_j$

$$d(x_i, x_j) := \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(\hat{y}^l_{ihw} - \hat{y}^l_{jhw}\right)\right\|_2^2$$

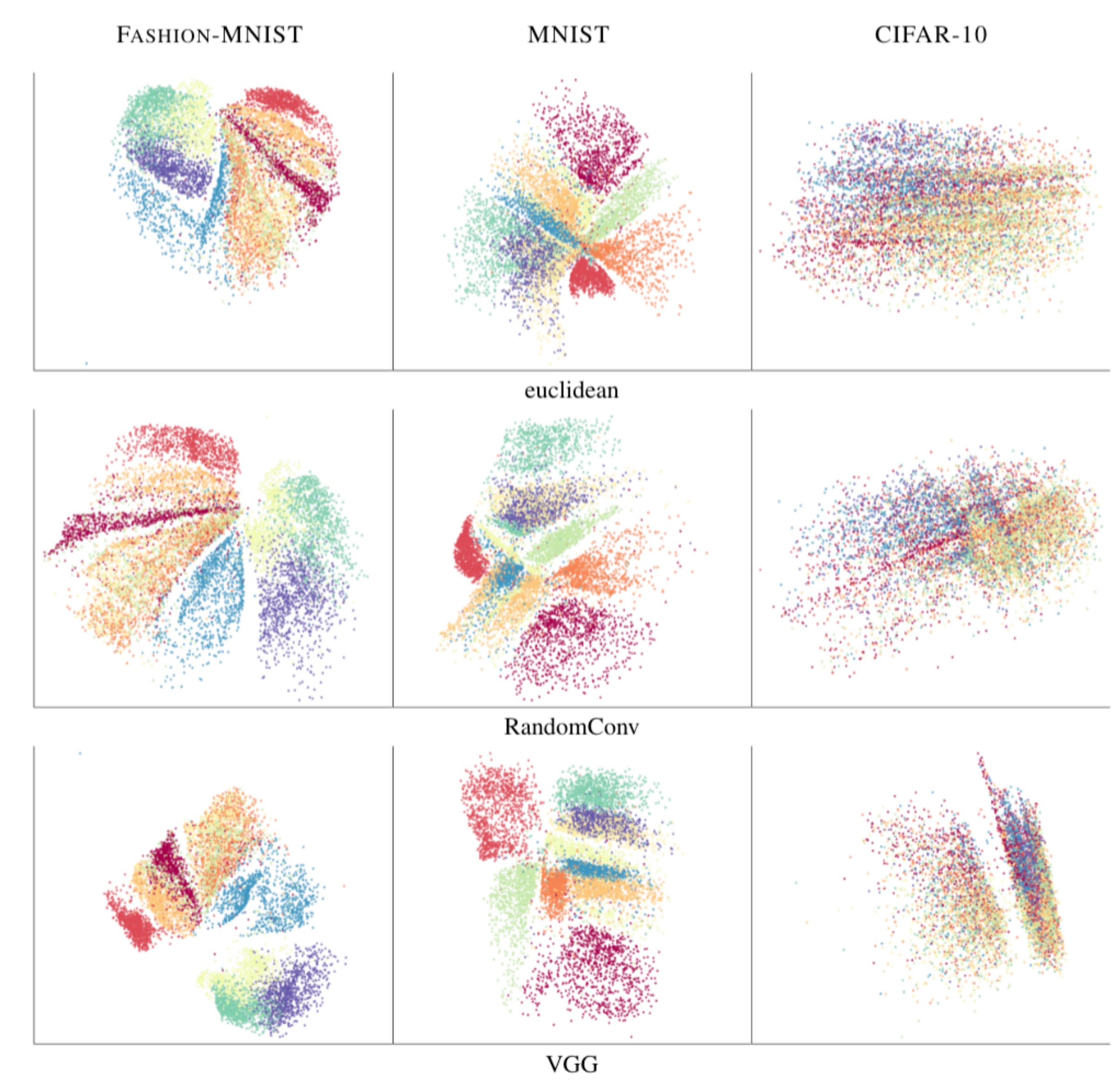
which corresponds to a per-channel scaled squared ℓ_2 distance (with w_l being the scale factor) of the features averaged over all spatial locations. Here, H_l and W_l denote the number of elements in the height direction and width direction, respectively.

In order to compute full distance matrices efficiently for a given pairwise LPIPS distance (with batch support), we contributed the PyTorch-compatible python package "batchdist" [3].

Experiments

Following [1] we use the datasets CIFAR-10, Fashion MNIST and MNIST, perform dimensionality reduction to two dimensions, and quantitatively evaluate the embeddings. TopoAE is referred to as "Euclidean", the random convolutions as "RandomConv" and the VGG-based LPIPS approach as "VGG". Furthermore, we visualise the latent encodings.

Dataset	Metric Model	$\mathrm{KL}_{0.01}$	$\mathrm{KL}_{0.1}$	KL_1	ℓ-Cont	ℓ-MRRE	ℓ-Trust	ℓ-RMSE	Data MSE
CIFAR	RandomConv	**0.573000**	0.026776	0.000519	0.884307	0.116578	**0.865882**	37.986596	**0.13567**
	VGG	0.747940	0.035598	0.000545	0.852621	0.134274	0.857791	38.790536	0.18757
	Euclidean	0.589208	**0.021210**	**0.000324**	**0.919586**	**0.107334**	0.851644	**37.628791**	0.14111
F-MNIST	RandomConv	0.421326	0.066843	0.001259	0.973635	0.025571	0.970463	22.048758	**0.10510**
	VGG	**0.380293**	**0.055905**	0.001052	0.977465	**0.023314**	**0.973091**	22.813647	0.10957
	Euclidean	0.391267	0.060878	**0.000981**	**0.980441**	0.025026	0.967511	**21.436654**	0.10919
MNIST	RandomConv	0.355687	0.130055	0.001506	0.921422	0.060333	0.930401	19.791551	0.14990
	VGG	0.674952	0.187962	0.001844	0.920575	**0.059505**	**0.931545**	20.174628	**0.14978**
	Euclidean	**0.343812**	**0.098164**	**0.000797**	**0.926969**	0.069833	0.906727	**18.976329**	0.15464



We found that the use of elaborate image distances did not result in marked improvements in terms of quality metrics, the Euclidean distance remains competitive. However, the resulting visualisations appear to exhibit an improved separation when employing alternative distances specific to the image domain.

Paper:

References

[1] Moor, M., Horn, M., Rieck, B., & Borgwardt, K. (2020, November). Topological autoencoders. In *Proceedings of the International Conference on Machine Learning* (pp. 7045-7054). PMLR.

[2] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 586-595).

[3] $ pip install batchdist

Code: