TOPOGEN: TOPOLOGY-INFORMED GENERATIVE MODELS

Jack Benarroch Jedlicki

Final Degree Thesis, Degree in Mathematics

Supervisors: Dr. Sergio Escalera, Dr. Carles Casacuberta

June, 2024

an anthropomorphic fractal person behind the counter at a fractal themed restaurant.

beautiful oil painting of a steamboat in a river in the afternoon. On the side of the river is a large brick building with a sign on top that says ŠD3˝.

an anthopomorphic pink donut with a mustache and cowboy hat standing by a log cabin in a forest with an old 1970s orange truck in the driveway

fox sitting in front of a computer in a messy room at night. On the screen is a 3d modeling program with a line render of a zebra.

Training generative models is hard!

Figure: Mode collapse in a generative adversarial network (GAN).

- DALLE2: 100k-200k GPU hours for training (more than 22 years with one GPU!).

Topology + Generative Models: a solution?

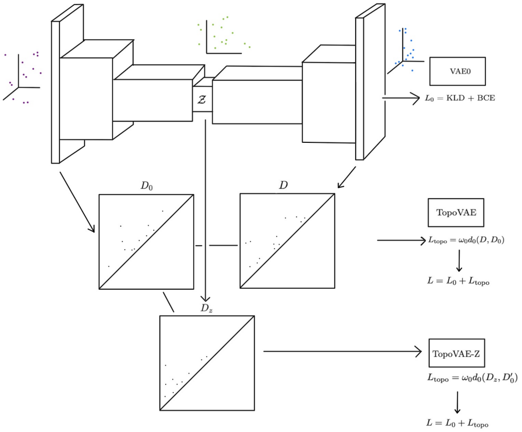
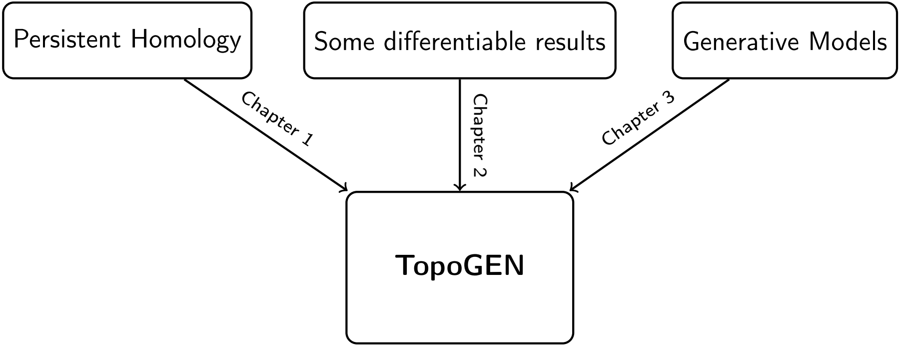


Figure: Working principle of a topology-informed variational autoencoder (TopoVAE).

Topology + Generative Models: a solution?



Index

1 Persistent homology

Definition: abstract simplicial complex

An *abstract simplicial complex* is a pair (K, S), where K is a set and S is a collection of subsets of K, such that for all $v \in K$ we have $\{v\} \in S$, and for all $\sigma \in S$, if $\tau \subseteq \sigma$ then $\tau \in S$.

- We name each element $\sigma \in S$ a k-simplex, where $k := |\sigma| - 1$ is its dimension, and a simplex of the form $\{v\}$ is called a vertex.
- Given a simplicial complex K and a k-simplex $\sigma = \{v_0, v_1, \ldots, v_k\}$, an orientation of σ is an equivalence class of orderings of the vertices of σ, where two orderings are equivalent if they differ by an even permutation. An oriented simplex is denoted as $\sigma = [v_0, v_1, \ldots, v_k]$.

Geometric simplicial complexes

The geometric simplicial complex K_G associated with K is a subspace of $\mathbb{R}^d$ formed by the convex hulls of all sets of points $\{v_{i_0}, v_{i_1}, \ldots, v_{i_k}\}$ where $[v_{i_0}, v_{i_1}, \ldots, v_{i_k}]$ is any k-simplex of K.

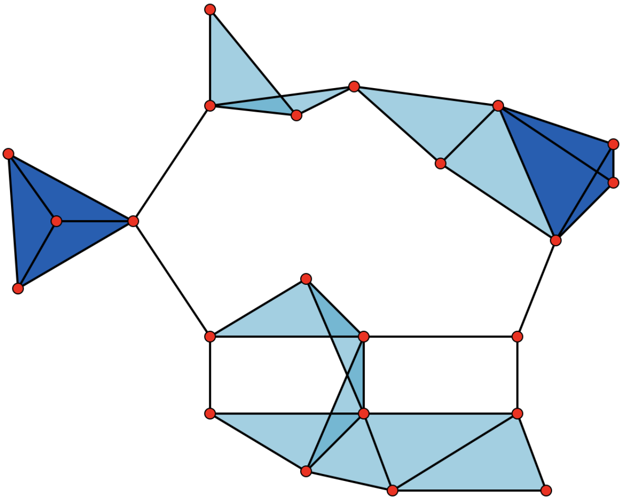


Figure: Example of a geometric simplicial complex.

Describing point clouds

Definition: point cloud

A *point cloud* (or (n, d)-*point cloud*) is an ordered set of n points $\{p_1, \ldots, p_n\}$ in $\mathbb{R}^d$ (for some positive integers n and d).

Definition: Rips complex

Let M be a metric space with metric d, and let $P = \{p_i\}_{i=1}^n \subseteq M$ be a point cloud in M. Given $\epsilon > 0$, the *Rips complex* $K_\epsilon(P)$ is the abstract simplicial complex with vertices the points of P, and given any ordered subset $\{i_0, \ldots, i_k\} \subseteq \{1, \ldots, n\}$, the k-simplex $[p_{i_0}, \ldots, p_{i_k}]$ exists in $K_\epsilon(P)$ if $d(p_x, p_y) \leq \epsilon$ for all $x, y \in \{i_0, \ldots, i_k\}$.

Definition: Rips filtration

The *Rips filtration* of P is the nested family of simplicial complexes $(K_\epsilon)_{\epsilon \in [0;+\infty)}$ where each K_ϵ is the Rips complex obtained from P with scale value ϵ.

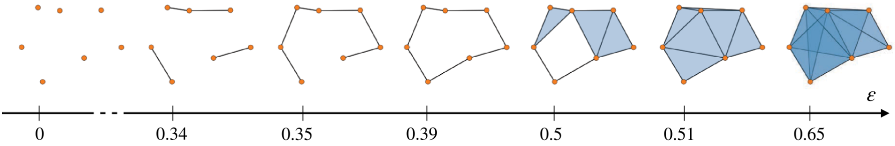
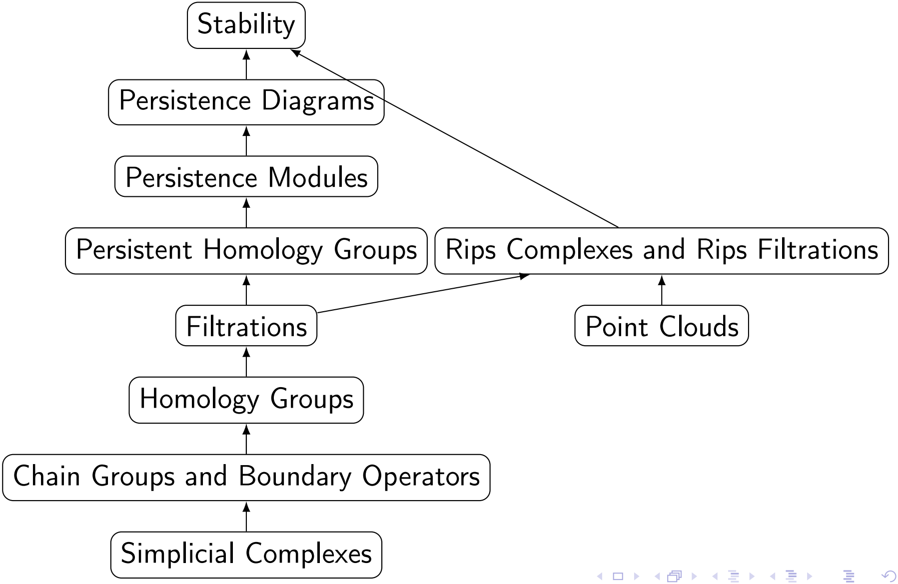


Figure: Example of simplicial complexes of the Rips filtration of a point cloud in the plane [1].

But how we can we trace the emergence and disappearance of features within the filtration?



But how we can we trace the emergence and disappearance of features within the filtration?

- The *persistence diagram* is a multi-set of intervals $(\epsilon_i, \epsilon'_i)$ where $\epsilon_i \in [0, +\infty)$ and $\epsilon'_i \in [0, +\infty]$, union Δ^∞, which can be computed from any point cloud.
- Each $(\epsilon_i, \epsilon'_i)$ corresponds to a k-hole with birth in K_{ϵ_i} and death in $K_{\epsilon'_i}$.
- The persistence diagram is a representation of the persistent homology (a family of vector spaces describing the "persistence" of features along the filtration).

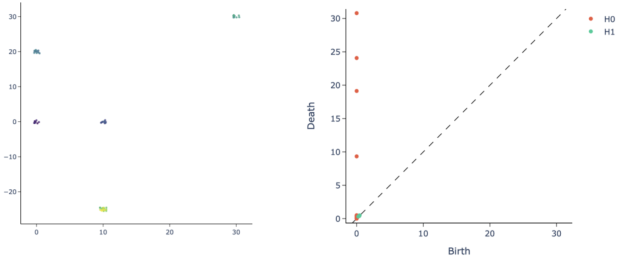


Figure: Example of a persistence diagram.

Definition

A *persistence diagram* or *barcode* is the union $B \cup \Delta^{\infty} \subseteq \mathbb{R} \times \bar{\mathbb{R}}$, where B is a finite multi-set of elements in $\mathbb{R} \times \bar{\mathbb{R}}$. The space Bar is the space of all persistence diagrams.

Definition: bottleneck distance

A *matching* between two barcodes D and D' is a bijection between the two multi-sets $\gamma \colon D \to D'$. The *cost* of a matching is defined as

$$c(\gamma) := \sup_{x \in D} \|x - \gamma(x)\|_{\infty},$$

where $\|\cdot\|_{\infty}$ is the supremum norm, given by $\|(a, b)\|_{\infty} = \max(|a|, |b|)$. We denote by $\Gamma(D, D')$ the set of all matchings between D and D'. The *bottleneck distance* between D and D' is then

$$d_{\infty}(D, D') := \inf_{\gamma \in \Gamma(D, D')} c(\gamma).$$

Stability of persistence diagrams

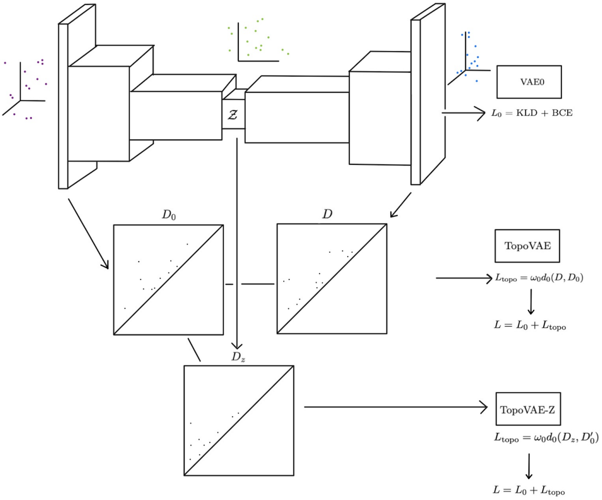
Given n and d, each point cloud $\{p_i\}_{i=1}^n \subseteq \mathbb{R}^d$ can be represented as a single point in $\mathbb{R}^{nd}$, and we refer to $\mathbb{R}^{nd}$ as the *space of point clouds*. More precisely:

$$\{p_i\}_{i=1}^n \subseteq \mathbb{R}^d \Leftrightarrow (p_1, \ldots, p_n) \in \mathbb{R}^{nd}.$$

Proposition

If $\|P - P'\|_2 < \epsilon$, *then* $d_\infty(\mathrm{Dgm}_k(f), \mathrm{Dgm}_k(f')) < 2\epsilon$, *where* $\mathrm{Dgm}_k(f)$ *and* $\mathrm{Dgm}_k(f')$ *are the persistence diagrams of degree* k *of* P *and* P' *via the Rips filtration, respectively.*

2 Bridging the gap between persistent homology and generative models



Definition: Topological regularizers

A *topological regularizer* is a function from the space of point clouds, $\mathbb{R}^{nd}$, to $\mathbb{R}$, factoring through Bar.

Gradient descent

Given a loss $L(\mathbf{a})$, where $\mathbf{a}$ is the vector of weights of the machine learning model, and a step size η, the weights are updated according to

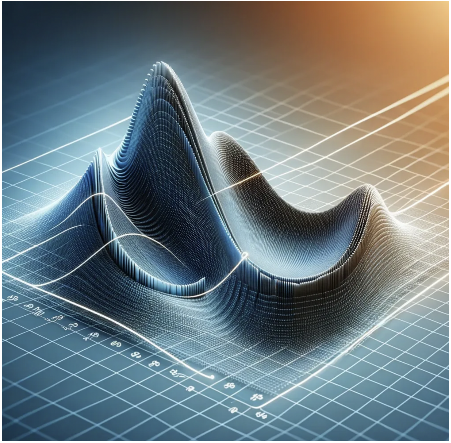
$$\mathbf{a}' = \mathbf{a} - \eta \nabla L(\mathbf{a}),$$

where $\nabla L(\mathbf{a})$ is the gradient of L, and $\mathbf{a}'$ are the updated weights.

(Due to the fact that if $F(\mathbf{x})$ is differentiable in a neighborhood of $\mathbf{x}_0$, then F decreases fastest in the direction of $-\nabla L(\mathbf{x}_0)$.)

Question

How do we know if a topological regularizer is differentiable?



2.1. Framework of differentiability through Bar

- Framework of differentiability developed by Leygonie et al. [2].
- Summary: Given $F\colon \mathcal{M} \xrightarrow{F_1} \mathsf{Bar} \xrightarrow{F_2} \mathcal{N}$, study the "differentiability" of $F_1\colon \mathcal{M} \to \mathsf{Bar}$ and of $F_2\colon \mathsf{Bar} \to \mathcal{N}$, and connect them using the "barcode chain rule".

Definition

Let $\mathbb{R}^{nd}$ be the space of point clouds. Given $p \in \mathbb{N}$, the *barcode generator of degree* p (via the Rips filtration) is the map

$$
\begin{aligned}
B_p\colon \mathbb{R}^{nd} &\to \mathsf{Bar} \\
P &\mapsto B_p(P),
\end{aligned}
$$

where $B_p(P)$ is the persistence diagram of degree p of P obtained via the Rips filtration.

2.2. Some differentiability results

2.2.1. Push functions

Definition

The *push function of degree p* is defined as

$$\begin{aligned} P_p^\circ \colon \mathbb{R}^{nd} &\to \mathbb{R} \\ P &\mapsto -\sum_{(b,d)\in B_p(P)} (d-b). \end{aligned} \tag{1}$$

Proposition

The push function P_p° of degree p is generically C^∞ in $\mathbb{R}^{nd}$.

2.2.2. Reininghaus dissimilarity

Definition

The persistence scale space kernel k_σ is given by [3]:

$$k_\sigma(F, G) = \frac{1}{8\pi\sigma} \sum_{p\in F,\, q\in G} e^{-\frac{\|p-q\|_2^2}{8\sigma}} - e^{-\frac{\|p-\bar{q}\|_2^2}{8\sigma}}, \tag{2}$$

where $\bar{q} = (q_2, q_1)$ for $q = (q_1, q_2)$, and the sum goes over all pairs $p \in F$, $q \in G$ of off-diagonal bounded points of both diagrams. This kernel induces a pseudo-distance d_σ in Bar, the *Reininghaus dissimilarity*, given by

$$d_\sigma(F, G) = \sqrt{k_\sigma(F, F) + k_\sigma(G, G) - 2k_\sigma(F, G)}. \tag{3}$$

Proposition

Let $\tilde{d}_{\sigma,D_0} : \mathbb{R}^{nd} \to \mathbb{R}$ *be the map defined as* $\tilde{d}_{\sigma,D_0}(P) = d_\sigma^2(B_p(P), D_0)$ *for any* $P \in \mathbb{R}^{nd}$. *Then,* $\tilde{d}_{\sigma,D_0}$ *is generically* C^∞ *in* $\mathbb{R}^{nd}$.

2.2.3. Scaled Gaussian density estimators

Definition: 4SGDEs

A $4SGDE$, denoted by E_4, is a function parametrized by $s > 0$ and $\sigma > 0$, given by

$$E_4(D, x; \sigma, s) = s \sum_{(b,d)\in D} (d-b)^4 e^{-((d-x)/\sigma)^2}.$$

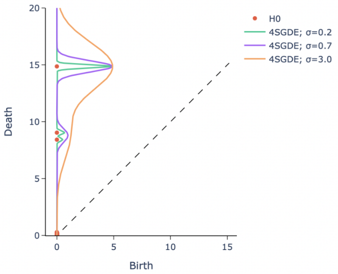


Figure: 4SGDE of a 0-dimensional persistence diagram, with $s = 0.0001$.

Proposition

Let E_4 be a 4SGDE, $D_0 \in \mathrm{Bar}$, and $\{x_1, \ldots, x_m\} \subseteq [0; +\infty)$. The function $L_E : \mathbb{R}^{nd} \to \mathbb{R}$ given by

$$L_{E_4}(P) = \frac{1}{m}\sum_{i=1}^{m}[E_4(B_0(P); x_i) - E_4(D_0; x_i)]^2$$

is generically C^∞ in $\mathbb{R}^{nd}$.

2.2.4. Persistent entropy

Definition

The *persistent entropy* of a persistence diagram D is defined as

$$\epsilon(D) = -\sum_{(b,d)\in D} \frac{|d-b|}{L} \log\left(\frac{|d-b|}{L}\right) \tag{4}$$

where $L = \sum_{(b,d)\in D} |d-b|$, and the summation only includes bounded off-diagonal points.

Problem: it is not differentiable in the diagonal.

Question

What about persistent entropy of diagrams of any degree p? Or more general classes of functions that are not differentiable in the diagonal, or that are not zero in the diagonal?

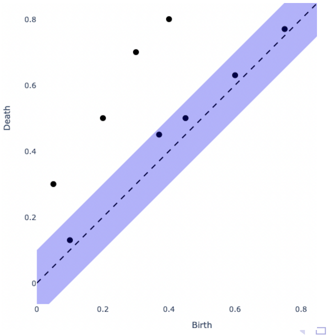
2.2.5. Selective regularizers

Definition: δ-selective functions

Given $\delta > 0$, a δ-*selective* function is a function $f_\delta \colon \mathrm{Bar}_0 \to \mathbb{R}$ of the form

$$f_\delta(D) = \sum_{(b,d)\in D\colon\, |d-b|>\delta} F(b,d) \tag{5}$$

where $F : \mathbb{R}^2 \to \mathbb{R}$ is C^∞ in $\{(x,y)\colon |x-y| > \epsilon\}$ for some $\epsilon < \delta$.



Definition

Given any $\delta > 0$, the *δ-selective $\mathbf{p}$-persistent entropy generator* $G_{p,\delta}$ is defined as

$$G_{p,\delta}\colon \mathbb{R}^{nd} \to \mathbb{R}$$

$$P \mapsto \epsilon(B_p(P) \setminus \bar{\Delta}_{\delta/2}) = -\sum_{(b,d)\in B_p(P)\colon\, |d-b|>\delta} \frac{|d-b|}{L} \log\left(\frac{|d-b|}{L}\right),$$

where $L = \sum_{(b,d)\in B_p(P)\colon\, |d-b|>\delta} |d-b|$, and $\bar{\Delta}_{\delta/2}$ is the set of points at a distance smaller or equal than $\delta/2$ to the diagonal.

$\rightarrow$ The map $G_{p,\delta}$ computes the persistent entropy of $B_p(P)$ "only looking at points in the diagram with persistence larger than δ".

Proposition

For any $p \in \mathbb{N}$ and $\delta > 0$, the δ-selective p-persistent entropy generator $G_{p,\delta}\colon \mathbb{R}^{nd} \to \mathbb{R}$ is generically C^∞ in $\mathbb{R}^{nd}$.

Definition

Let $\mathbb{R}^{nd}$ be the space of point clouds, $D_0 \in \mathsf{Bar}$ a fixed barcode, and $p \geq 0$. We have the five following types of topological regularizers, from $\mathbb{R}^{nd}$ to $\mathbb{R}$:

$$
\begin{aligned}
L_{\mathrm{push},p} \colon P &\mapsto -\sum_{(b,d)\in B_p(P)} (d-b), \\
L_{\mathrm{Rh},p}(\sigma) \colon P &\mapsto d_\sigma^2(B_p(P), D_0), \\
L_{\mathrm{4SGDE},0}(\mathsf{x}, \sigma_d, s) \colon P &\mapsto \sum_{i=1}^{m} [E_4(B_0(P); x_i, \sigma_d, s) - E_4(D_0; x_i, \sigma_d, s)]^2, \\
L_{\mathrm{bottleneck},p} \colon P &\mapsto d_\infty(B_p(P), D_0), \\
L_{\mathrm{entropy},p}(\delta) \colon P &\mapsto \left(G_{p,\delta}(P) - G_{p,\delta}(D_0)\right)^2 .
\end{aligned}
$$

We refer to these functions, respectively, as the *p-push regularizer*, the *p-Reininghaus regularizer*, the *4SGDE regularizer*, the *p-bottleneck regularizer*, and the *p-entropy regularizer*.

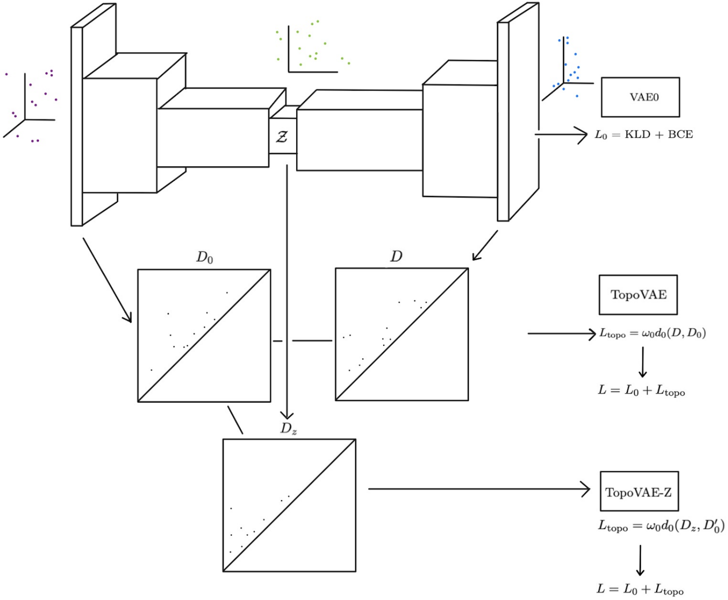
Theorem

Let $\mathbb{R}^{nd}$ be the space of point clouds. Then, the p-push regularizers, p-Reininghaus regularizers, 4SGDE *regularizers, and p-entropy regularizers are generically differentiable in $\mathbb{R}^{nd}$. In addition, the p-bottleneck regularizers are differentiable whenever the point cloud is in a generic subspace of $\mathbb{R}^{nd}$, and its persistence diagram satisfies mild computational conditions.*

Next step: testing them

Topology-informed generative models

3.1. Topology-informed variational autoencoders



Four topology-informed total losses

$$L_{T1}(P;\omega_0,\omega_1) = L_0 + \omega_0 \cdot L_{\text{bottleneck},0}(P;D_0^0) + \omega_1 \cdot L_{\text{bottleneck},1}(P;D_0^1),$$
$$L_{T2}(P;\omega_0,\omega_1,\delta) = L_0 + \omega_0 \cdot L_{\text{entropy},0}(P;D_0^0,\delta) + \omega_1 \cdot L_{\text{entropy},1}(P;D_0^1,\delta),$$
$$L_{T3}(P;\omega_0,\omega_1,\delta,\sigma) = L_0 + \omega_0 \cdot L_{\text{entropy},0}(P;D_0^0,\delta) + \omega_1 \cdot L_{\text{Rh},1}(P;D_0^1,\sigma),$$
$$L_{T4}(P;\omega_0,\omega_1,\mathbf{x},\sigma_d,s,\sigma) = L_0 + \omega_0 \cdot L_{\text{4SGDE},0}(P;D_0^0,\mathbf{x},\sigma_d,s)$$
$$+\,\omega_1 \cdot L_{\text{Rh},1}(P;D_0^1,\sigma),$$

where $L_0 = \text{KLD} + \text{BCE}$ (standard loss of a VAE).

3.2. Experiments and results

- VAE used: 1.68M parameters; latent space dimension: 10.
- Dataset: FashionMNIST (60k images, 10 image classes).
- Qualitative comparison: quality and diversity of generated images, changes in the decay of BCE losses (1 epoch training).

TopoVAE1 (bottleneck regularizers)

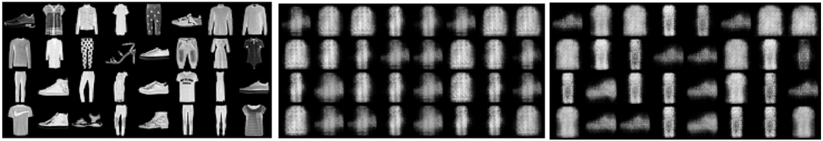


Figure: Comparison of TopoVAE1 and VAE0 at iteration 50/469, using $(\omega_0, \omega_1) = (5.0, 5.0)$.

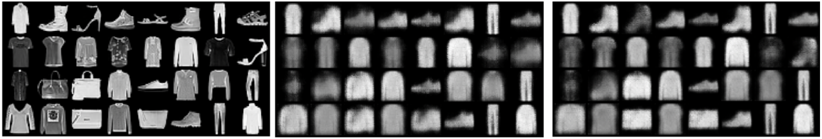


Figure: Comparison of TopoVAE1 and VAE0 at iteration 100/469, using $(\omega_0, \omega_1) = (5.0, 5.0)$.

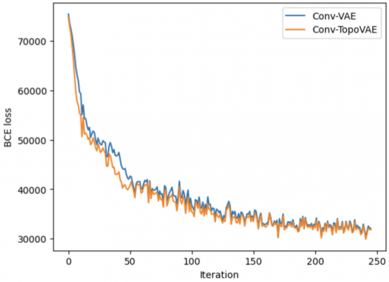


Figure: Evolution of the BCE lossses of TopoVAE1 and VAE0 used in the previous figure.

TopoVAE2 (entropy regularizers)

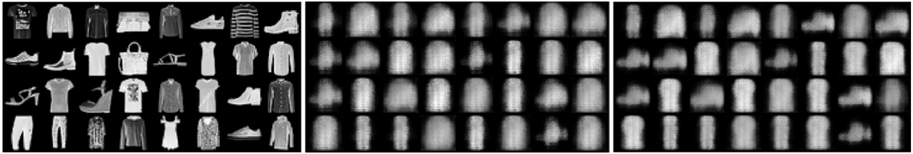


Figure: Comparison of TopoVAE2 and VAE0 using initially cloned models. Iteration 50/469.

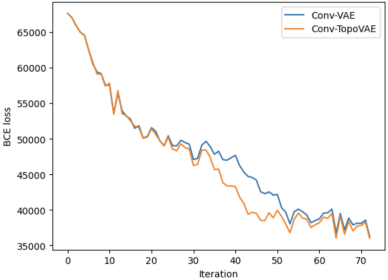


Figure: Evolution of the BCE lossses of TopoVAE2 and VAE0.

TopoVAE3: 0-entropy regularizer + 1-Reininghaus regularizer

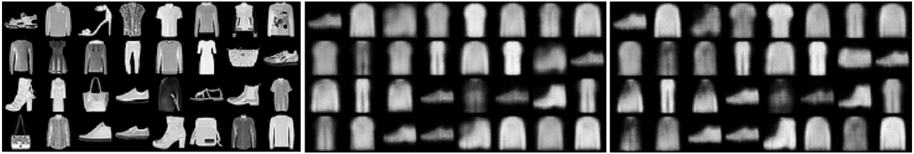


Figure: Comparison of TopoVAE3 and VAE0, iteration 125/469.

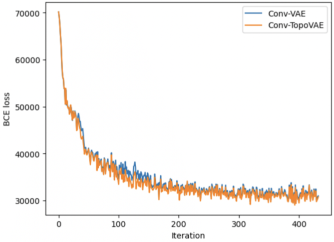


Figure: Evolution of the BCE lossses of TopoVAE3 ("Conv-TopoVAE") and VAE0 ("Conv-VAE"), used in the previous figure.

TopoVAE4: 4SGDE + 1-Reininghaus regularizer

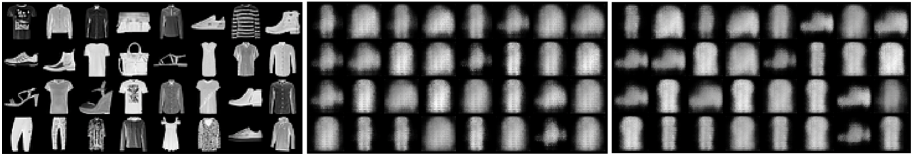


Figure: Comparison of TopoVAE4 and VAE0, iteration 50/469.

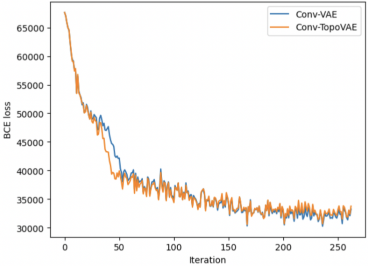


Figure: Evolution of the BCE lossses of TopoVAE4 and VAE0.

Quantitative comparison: results

- Quantitative comparison: computation of $r(s)$ values, where $r(s)$ is

$$r(s) = 100 \cdot \frac{\text{BCE}_{\text{VAE0}}(s) - \text{BCE}_{\text{TopoVAE}i}(s)}{\text{BCE}_{\text{VAE0}}(s)},$$

for $s = 25, 50, 75$, and computation of the means $\bar{r}(s)$ for each model.

Model	**Time**	$\bar{r}(25)$	$\bar{r}(50)$	$\bar{r}(75)$
VAE0	1.5 sec.			
TopoVAE1	7.1 sec.	0.3 %	1.2 %	0.2 %
TopoVAE2	2.1 sec.	0.2 %	1.1 %	0.3 %
TopoVAE3	2.1 sec.	0.2 %	0.7 %	0.2 %
TopoVAE4	17.0 sec.	0.2 %	0.6 %	0.2%

Table 1. Running time of 10 training steps, and $\bar{r}$ values at training steps 25, 50, and 75, for each TopoVAE.

3.3. Opening new lines of exploration: regularization of latent space

TopoVAE-Z1

The total loss of TopoVAE-Z1 is

$$L_{T1}(P;\omega_0,\omega_1) = L_0 + \omega_0 \cdot L_{\text{bottleneck},0}(P; D_0^{0\prime}) + \omega_1 \cdot L_{\text{bottleneck},1}(P; D_0^{1\prime}),$$

where P is the batch of latent vectors (and not the batch of final outputs of the model), and $D_0^{0\prime}$ and $D_0^{1\prime}$ are rescaled versions of the persistence diagrams D_0^0 and D_0^1 from the input batch of ground truth images.

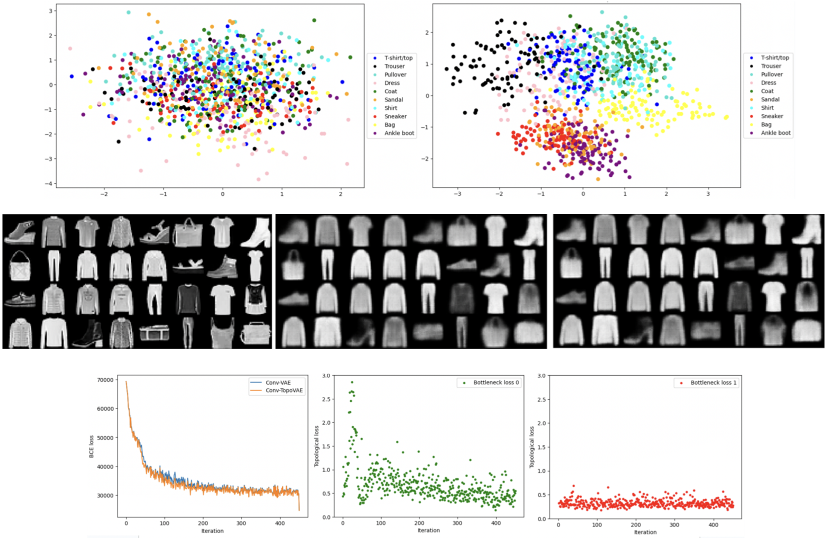


Figure: Comparison of VAE0 and TopoVAE-Z1. Top: distribution of 1000 latent vectors after 1 epoch, middle: images produced at iteration 125; bottom: comparison of losses during 1 training epoch.

Conclusions

- Development of a new approach for training generative models, merging them with persistent homology.
- New differentiability results, expanding the current knowledge of differentiability through barcode space to more general classes of functions.
- Experimental results showing that topological regularization can enhance the learning process of generative models.
- Future work: currently working on topology+diffusion (at the Computer Vision Center).

Questions

Bibliography

[1] Marco Guerra et al. "Homological scaffold via minimal homology bases". In: *Scientific reports* 11.1 (2021), p. 5355.

[2] Jacob Leygonie, Steve Oudot, and Ulrike Tillmann. "A framework for differential calculus on persistence barcodes". In: *Foundations of Computational Mathematics* (2021), pp. 1–63.

[3] Jan Reininghaus et al. "A stable multi-scale kernel for topological machine learning". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 4741–4748.