Persistent Topology for Natural Data Analysis — A Survey

Massimo Ferri[✉]

Dip. di Matematica and ARCES, Univ. di Bologna, Bologna, Italy
massimo.ferri@unibo.it

Abstract. Natural data offer a hard challenge to data analysis. One set of tools is being developed by several teams to face this difficult task: Persistent topology. After a brief introduction to this theory, some applications to the analysis and classification of cells, liver and skin lesions, music pieces, gait, oil and gas reservoirs, cyclones, galaxies, bones, brain connections, languages, handwritten and gestured letters are shown.

Keywords: Homology · Betti numbers · Size functions · Filtering function · Classification · Retrieval

1 Introduction and Motivation

What is the particular challenge offered by natural data, which could suggest the need of topology, and in particular of persistence? Simply said, it's quality instead of quantity. This is especially evident with images.

If one has to analyze, classify, retrieve images of mechanical pieces, vehicles, rigid objects, then geometry fulfills all needs. On the images themselves, matrix theory provides the transformations for superimposing a picture to a template. More often, pictures are represented by feature vectors, whose components are geometric measures (*shape descriptors*). Then recognition, defect detection, retrieval etc. can be performed on the feature vectors.

The scene changes if the depicted objects are of natural origin: the rigidity of geometry becomes an obstacle. Recognizing the resemblance between a sitting and a standing man is difficult. The challenge is even harder when it comes to biomedical data and when the context is essential for the understanding of data [34,51].

It's here that topology comes into play: the standing and sitting men are *homeomorphic*, i.e. there is a topological transformation which superimposes one to the other, whereas no matrix will ever be able to do that. It is generally difficult to discover whether two objects are homeomorphic; then algebraic topology turns helpful: It associates invariants — e.g. Betti numbers — to topological spaces, such that objects which are homeomorphic have identical invariants (the converse does not hold, unfortunately).

(Algebraic) topology seems then to be the right environment for formalizing qualitative aspects in a computable way, as is nicely expressed in [35, Sect. 5.1].

There is a problem: if geometry is too rigid, topology is too free. This is the reason why persistent topology can offer new topological descriptors (e.g. Persistent Betti Numbers, Persistence Diagrams) which preserve some selected geometric features through *filtering functions*. Classical references on persistence are [8, 12, 22, 50].

Persistent topology has been experimented in the image context, particularly in the biomedical domain, but also in fields where data are not pictures, e.g. in geology, music and linguistics, as will be shown in this survey.

2   Glossary and Basic Notions

It is out of the scope of this survey to give a working introduction to homology and persistence; we limit ourselves to an intuitive description of the concepts, and recommend to profit of the technical references, without which a real understanding of the results is impossible. An essential (and avoidable) technical description of a particular homology is reported in Sect. 2.1.

Homology. There is a well-structured way (technically a set of functors) to associate *homology vector spaces* (more generally modules) $H_k(X)$ to a simplicial complex or to a topological space X, and linear transformations to maps [33, Chap. 2] and [23, Chap. 4].

Betti numbers. The *k-th Betti number* $\beta_k(X)$ is the dimension of the k-th homology vector space $H_k(X)$, i.e. the number of independent generators (homology classes of *k-cycles*) of this space. Intuitively, $\beta_0(X)$ counts the number of path-connected components (i.e. the separate pieces) of which X is composed; $\beta_1(X)$ counts the holes of the type of a circle (like the one of a doughnut); $\beta_2(X)$ counts the 2-dimensional voids (like the ones of gruyere or of an air chamber).

Homeomorphism. Given topological spaces X and Y, a *homeomorphism* from X to Y is a continuous map with continuous inverse. If one exists, the two spaces are said to be *homeomorphic*. This is the typical equivalence relation between topological spaces. Homology vector spaces and Betti numbers are invariant under homeomorphisms.

Remark 1. As hinted in the Introduction, geometry is too rigid, but topology is too free. In particular, homeomorphic spaces can be very different from an intuitive viewpoint: the joke by which "for a topologist a mug and a doughnut are the same" is actually true; the two objects are homeomorphic! **Persistent topology** then tries to overcome this difficulty by studying not just topological spaces but pairs, once called *size pairs*, (X, f) where f is generally a continuous function, called *measuring* or *filtering function*, from X to $\mathbb{R}$ (to $\mathbb{R}^n$ in *multidimensional persistence*) which conveys the idea of shape, the viewpoint of the observer. Shape similarity is actually very much dependent on the context. The Betti numbers of the sublevel sets then make it possible to distinguish the two objects although they are homeomorphic: see Fig. 1.
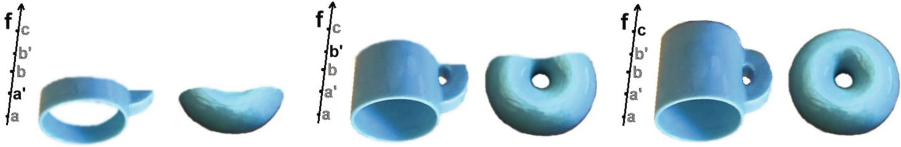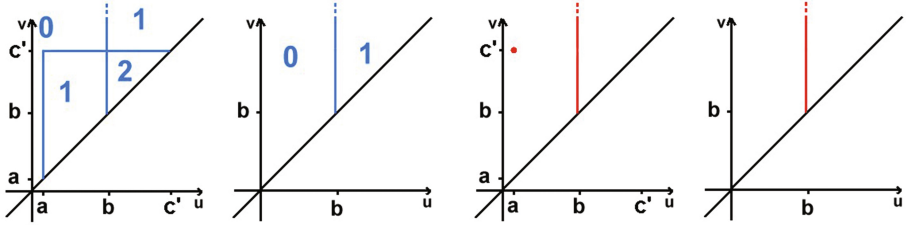
Fig. 1. Sublevel sets of mug and doughnut.

Fig. 2. From left to right: 1-PBN functions of mug and of doughnut, 1-PDs of mug and of doughnut.

Sublevel Sets. Given a pair (X, f), with $f : X \to \mathbb{R}$ continuous, given $u \in \mathbb{R}$, the *sublevel set under* u is the set $X_u = \{x \in X \mid f(x) \leq u\}$.

Persistent Betti Numbers. For all $u, v \in \mathbb{R}$, $\quad u < v$, the inclusion map $\iota^{u,v} : X_u \to X_v$ is continuous and induces, at each degree k, a linear transformation $\iota_*^{u,v} : H_k(X_u) \to H_k(X_v)$. The *k-Persistent Betti Number (k-PBN) function* assigns to the pair (u, v) the number $\dim \operatorname{Im} \iota_*^{u,v}$, i.e. the number of classes of k-cycles of $H_k(X_u)$ which "survive" in $H_k(X_v)$. See Fig. 2 (left) for the 1-PBN functions of mug and doughnut. Note that a pitcher, and more generally any open container with a handle, will have very similar PBNs to the ones of the mug; this is precisely what we want for a functional search and not for a strictly geometrical one.

Persistence Diagrams. The k-PBN functions are wholly determined by the position of some discontinuity points and lines, called *cornerpoints* and *cornerlines* (or *cornerpoints at infinity*) The coordinates (u, v) of a cornerpoint represent the levels of "birth" and "death" respectively of a generator; the abscissa of a cornerline is the level of birth of a generator which never dies. The *persistence* of a cornerpoint is the difference $v - u$ of its coordinates. Cornerpoints and cornerlines form the *k-Persistence Diagram (k-PD)*. Figure 2 (right) depicts the 1-PDs of mug and of doughnut. For the sake of simplicity, we are here neglecting the fact that cornerpoints and cornerlines may have multiplicities.

Remark 2. Sometimes it is important to distinguish even objects for which there exists a rigid movement superimposing one to the other — so also geometrically equivalent — as in the case of some letters: context may be essential! See Fig. 3, where ordinate plays the role of filtering function.
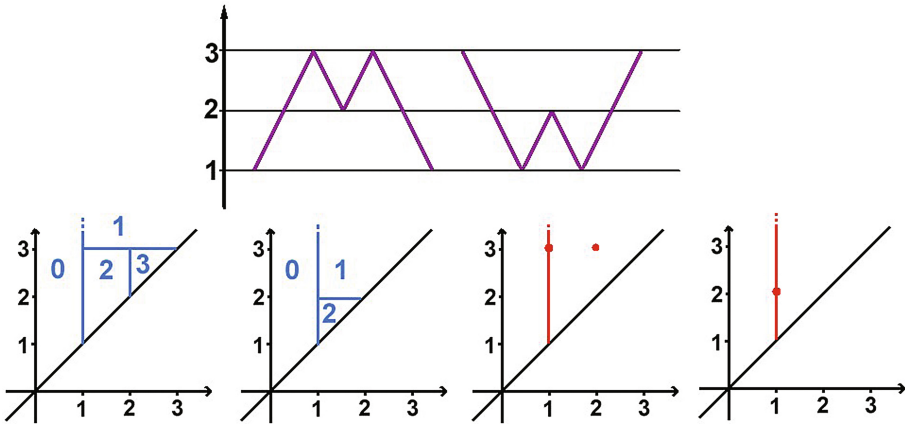
Fig. 3. Above: the objects "M" and "W". Below, from left to right: 0-PBN functions of M and of W, 0-PDs of M and of W.

Matching distance. Given the k-PDs $\mathcal{D}_{X,f}, \mathcal{D}_{Y,g}$ of two pairs $(X, f), (Y, g)$, match the cornerpoints of $\mathcal{D}_{X,f}$ either with cornerpoints of $\mathcal{D}_{Y,g}$ or with their own projections on the diagonal $u = v$; the *weight* of this matching is the sup of the L_∞-distances of matching points. The *matching distance* (or *bottleneck distance*) of $\mathcal{D}_{X,f}$ and $\mathcal{D}_{Y,g}$ is the inf of such weights among all possible such matchings.

Natural pseudodistance. Given two pairs $(X, f), (Y, g)$, with X, Y homeomorphic, the *weight* of a given homeomorphism $\varphi : X \to Y$ is $\sup_{x \in X} |g(\varphi(x)) - f(x)|$. The *natural pseudodistance* of (X, f) and (Y, g) is the inf of these weights among all possible homeomorphisms. If we are given the k-PDs of the two pairs, their matching distance is a lower bound for the natural pseudodistance of the two pairs, and it is the best possible obtainable from the two k-PDs. Much is known on this dissimilarity measure [19–21].

2.1 A Brief Technical Description of Homology

There are several homologies. The classical and most descriptive one, at least for compact spaces, is singular homology with coefficients in $\mathbb{Z}$; we refer to [33, Chap. 2] for a thorough exposition of it. Anyway, the homology used in most applications is the simplicial one, of which (with coefficients in $\mathbb{Z}_2$) we now give a very short introduction following [23, Chap. 4].

Simplices. A *p-simplex* σ is the convex hull, in a Euclidean space, of a set of $p+1$ points, called *vertices* of the simplex, not contained in a Euclidean $(p-1)$-dimensional subspace; the simplex is said to be *generated* by its vertices. A *face* of a simplex σ is the simplex generated by a nonempty set of vertices of σ.

Simplicial complexes. A finite collection K of simplices of a given Euclidean space is a *simplicial complex* if (1) for any $\sigma \in K$, all faces of σ belong to K,
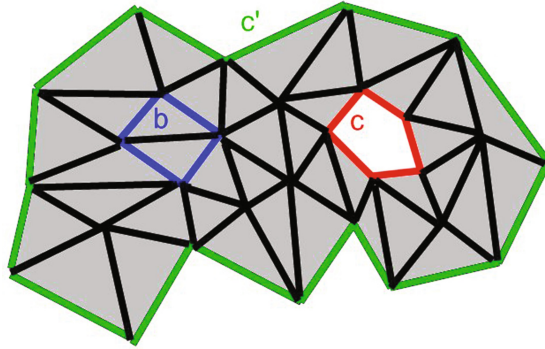
Fig. 4. Cycles.

(2) the intersection of two simplices of K is either empty or a common face. The *space* of the complex K is the topological subspace of Euclidean space $|K|$ formed by the union of all simplices of K.

Simplicial homology with $\mathbb{Z}_2$ coefficients. Given a (finite) simplicial complex K, call *p-chain* any formal linear combination of p-simplices with coefficients in $\mathbb{Z}_2$ (i.e. either 1 or 0, with $1 + 1 = 0$). p-chains form a $\mathbb{Z}_2$-vector space C_p. Note that each p-chain actually identifies a set of p-simplices of K and that the sum of two p-chains is just the symmetric difference (Xor) of the corresponding sets. We now introduce a linear transformation $\partial_p : C_p \to C_{p-1}$ (called *boundary operator*) for any $p \in \mathbb{Z}$. We just need to define it on generators, i.e. on p-simplices, and then extend by linearity. Writing $\sigma = [u_0, u_1, \ldots, u_p]$, we denote by $[u_0, \ldots, \hat{u}_j, \ldots, u_p]$ its face generated by all of its vertices except u_j $(j = 0, \ldots, p)$. Then we define

$$\partial_p(\sigma) = \sum_{j=0}^{n} [u_0, \ldots, \hat{u}_j, \ldots, u_p]$$

It is possible to prove that $\partial_p \partial_{p+1} = 0$, so that $B_p = \mathrm{Im}\partial_{p+1}$ is contained in $Z_p = \mathrm{Ker}\partial_p$. Elements of B_p are called *p-boundaries*; elements of Z_p are called *p-cycles*. The *p-homology vector space* is defined as the quotient $H_p(K) = Z_p/B_p$. Homology classes are represented by cycles which are not boundaries. Two cycles are *homologous* is their difference is a boundary. In Fig. 4, representing the simplicial complex K formed by the shaded triangles and their faces, the blue chain b is a 1-cycle which is also a boundary; the red chain c and the green one c' are 1-cycles which are not boundaries; c and c' are homologous.

3   State-of-the-Art

The application of persistence to shape analysis and classification has a long story, since it started in the 90's when it still had the name of Size Theory [50]. In the last few years it has taken various, very interesting forms. The constant
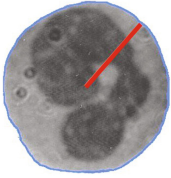
Non-basophile granulocyte   Basophile granulocyte

Lymphocyte   Monocyte

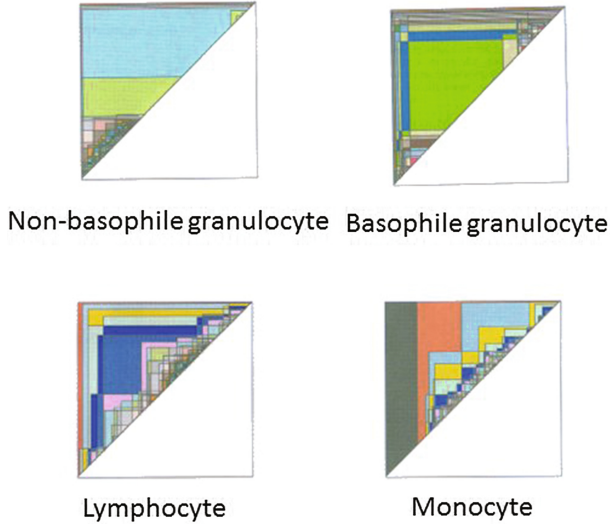**Fig. 5.** A radius along which the three filtering functions are computed.

Fig. 6. Persistent Betti Number functions relative to the sum of grey tones (different colors represent different values).

aspect is always the presence of qualitative features which are difficult to capture and formalize within other frames of mind.

3.1   Leukocytes

Leukocytes, or white blood cells, belong to five different classes: lymphocyte; monocyte; neutrophile, eosinophile, basophile granulocytes. Eosinophile and neutrophile granulocytes are generally difficult to be distinguished, so they were considered in a single classification class in an early research by the Bologna team [26].

As a space, the boundary of the starlike hull of the cell is assumed. The images are converted to grey tones.

Three filtering functions are put to work, all computed along radii from the center of mass of the cell (Fig. 5):

– Sum of grey tones
– Maximum variation
– Sum of variations pixel to pixel.

Classification (with very good hit ratios for that time) is performed by measuring distance from the average PBN function of each class.

3.2   Handwritten Letters and Monograms

Again in Bologna we faced recognition of handwritten letters with time information; our goal was to recognize both the alphabet letter and the writer [25].
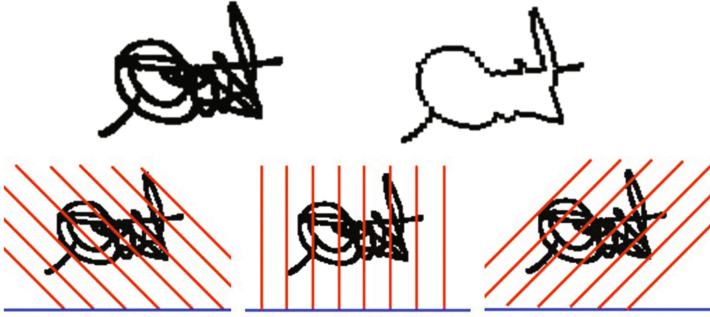
Fig. 7. A monogram with its outline (above) and the directions along which the filtering functions are computed (below).

The space on which the filtering functions are defined is the time interval of the writing. The filtering functions are computed in the 3D "plane-time":

- Distance of points from the letter axis
- Speed
- Curvature
- Torsion
- Distance from center of mass (in plane projection).

Classification comes from fuzzy characteristic functions, obtained from normalized inverse of distance. Cooperation of the characteristic functions coming from the single filtering functions is given by their rough arithmetic average.

A later experiment, which was even repeated live at a conference, concerned the recognition of monograms for personal identification, without time information [24].

Two topological spaces are used. The first is the outline of the monogram and the filtering function is the distance from the center of mass (see upper Fig. 7).

The second space is a horizontal segment placed at the base of the monogram image. Filtering functions:

- Number of black pixels along segments (3 directions) (see lower Fig. 7)
- Number of pixel-pixel black-white jumps (3 directions).

Classification is performed by a weighted average of fuzzy characteristic functions.

3.3   Sign Alphabet

Automatic recognition of the symbols expressed by the hands in the sign language is a task which was of interest for different teams. The first one was the group led by Alessandro Verri in Genova [49]. The signs were performed with a
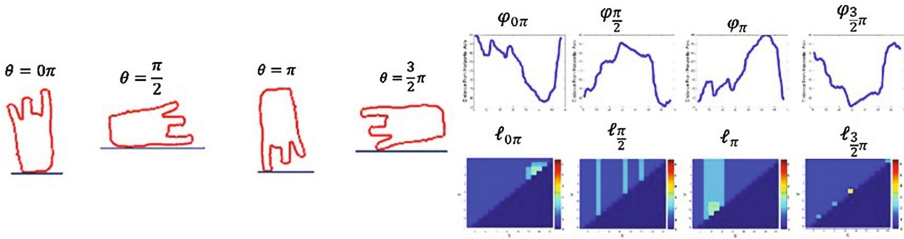
Fig. 8. Four filtering functions and the corresponding 0-Persistent Betti Number functions.
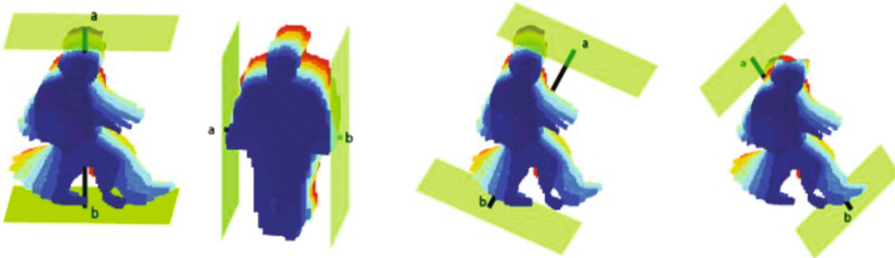

Fig. 9. Four filtering functions on silhouette stacks for gait identification.

white glove on a black background; translation into common letters was done in real time in a live demo at a conference.

The domain space is a horizontal segment; the filtering functions assign to each point of the segment the maximum distance of a contour point within a strip of fixed width, with 24 different strip orientations.

The choice of S. Wang in Sherbrooke, instead, is to use a part of the contour, determined by principal component analysis, as a domain and distance from center of mass as filtering function [32].

The team of D.Kelly in Maynooth uses the whole contour as domain, and distances from four lines as filtering functions [36] (see Fig. 8).

3.4 Human Gait

Personal identification and surveillance are the aim of a research by the Cuban team of L. Lamar-León, together with the Sevilla group of computational topology [37].

Considering a stack of silhouettes as a 3D object, and using four different filtering functions, makes 0- and 1-degree persistent homology a tool for identifying people through their gait (Fig. 9).
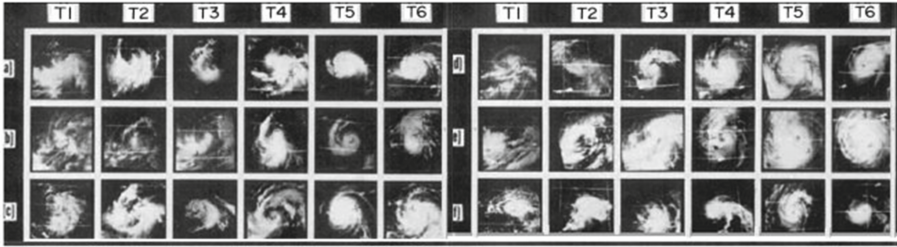
Fig. 10. Time evolution of cyclones.

3.5 Tropical Cyclones

S. Banerjee in Kolkata makes use of persistence on sequences of satellite images of cloud systems (Fig. 10), in order to evaluate risk and intensity of forming hurricanes [2].

Time interval is the domain of two filtering functions which are common characteristic measures of cyclones:

– Central Feature portion
– Outer Banding Feature

3.6 Galaxies

Again S. Banerjee [3] applies similar methods to another type of spirals: galaxies.

Various filtering functions are used. One is defined as a function of distance from galaxy center, and is the ratio between major and minor axis of the corresponding isophote. Another one is a "pitch" parameter defined by Ringermacher and Mead [45]. A third filtering function is a compound based on color.

The classification results agree with the literature.

3.7 Bones

In [48] a powerful construction (the Persistent Homology Transform) is introduced. It consists in gathering the "height" filtering functions according to all possible directions. The paper shows that the transform is injective for objects homeomorphic to spheres. By using the transform it is possible to define an effective distance between surfaces. An application is shown by classifying heel bones of different species; the comparison with the ground truth produced by using placement of landmarks on the surfaces is very good.

3.8 Melanocytic Lesions

A very important part of natural shape analysis is the detection of malignant cells and lesions, since there generally are no templates for them. As far as we
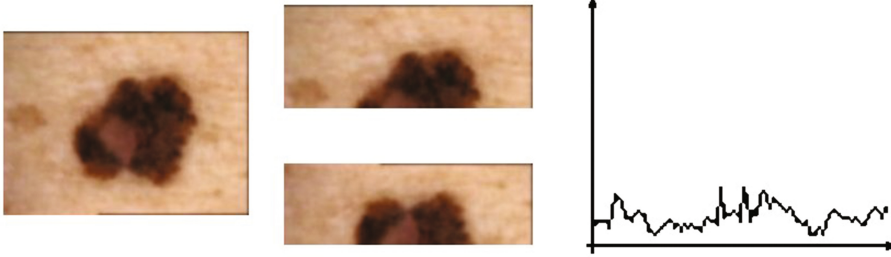
Fig. 11. One of the 45 splittings of a melanocytic lesion, and the whole *A*-curve corresponding to the filtering function luminance.

know, the first attempt through persistence (called *size theory* at that time) is the ADAM EU Project, by the Bologna team together with CINECA and with I. Stanganelli, a dermatologist of the Romagna Oncology Institute [17,27,47]. The analysis is mainly based on asymmetry of boundary, masses and color distribution: the lesion is split into two halves by 45 equally spaced lines, and the difference between the two halves is measured by the matching distance of the corresponding Persistence Diagrams.

The three functions (*A*-curves) relating these distances to the splitting line angles give parameters which are then fed into a Support Vector Machine classifier.

The same team is presently involved with a biomedical firm in the realization of a machine for smart retrieval of dermatological images [28].

3.9   Tumor Mouth Cells

A morphological classification of normal and tumor cells of the epithelial tissue of the mouth is proposed in [40,41]: the filtering function is distance from the center of mass; the discrimination is statistically based on the distribution of cornerpoints (see Fig. 12).

3.10   Hepatic Lesions

The advantages of a multidimensional range for the filtering functions are shown in [1], where several classification experiments are performed on the images of hepatic cells (see Fig. 13). The domain space is the part of image occupied by the lesion; the two components of the filtering function are the greyscale of each pixel and the distance from the lesion boundary.

3.11   Genetic Pathways

So far we have seen applications of persistence to images of natural origin. But the modularity of the method opens the possibility to deal with data of very different nature. A first example is given by [43], where persistence is used on the
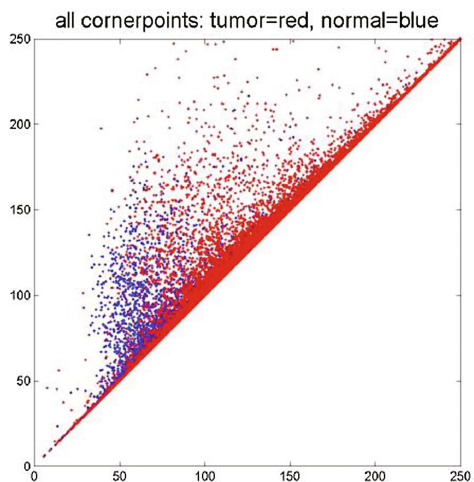
Fig. 12. Distribution of cornerpoints in the diagrams of normal and tumor mouth cells.

(a) Cyst

(b) Metastasis

(c) Hemangioma

(d) HCC

(e) Focal Nodule

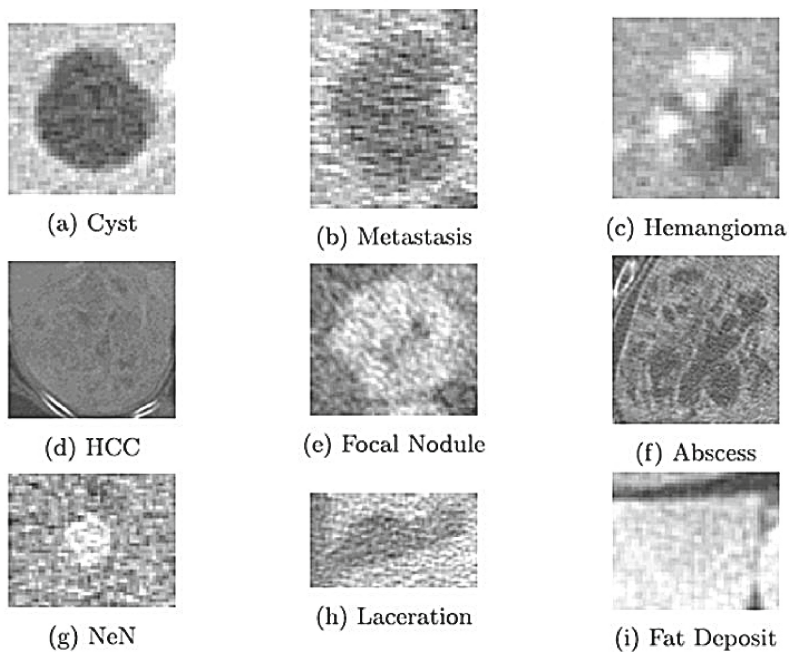(f) Abscess

(g) NeN

(h) Laceration

(i) Fat Deposit

Fig. 13. Various types of hepatic lesions.

Vietoris-Rips complex in a space where points are *complex phenotypes* related together by the *Jaccard distance*. This made it possible to find systematic associations among metabolic syndrome variates that show distinctive genetic association profiles.

3.12   Oil and Gas Reservoirs

Researchers in Ufa and Novosibirsk need to get a reliable geological and hydrodynamical model of gas and oil reservoirs out of noisy data; the model has to be robust under small perturbations. The authors have found an answer in persistent 0-, 1- and 2-cycles. The domain space is the 3D reservoir bed, and the filtering function is permeability, obtained as a decreasing function of radioactivity [4] (Russian; translated and completed in this same volume).

3.13   Brain Connections

A complex research on brain connections and their modification under the assumption of a psychoactive substance (psilocybine) is performed in [42] and extended in [39]. The construction starts with a complete graph whose vertices are cortical or subcortical regions; these, and their functional connectivity (expressed as weights on the edges) come from an elaborate processing of functional MRI data. Then the simplicial complex is built, whose simplices are the cliques (complete subgraphs) of the graph.

The filtering function on each simplex is minus the highest weight of its building edges. A difference between treated and control subjects already appears in the comparison of the 1-Persistence Diagrams (see Fig. 14). Then more information is obtained from secondary graphs (called *homological scaffolds*), whose vertices are the homology generators weighted by their persistence.

There are other applications of persistence to brain research: evaluation of cortical thickness in autism [16]; study of unexpected connections between subcortex, frontal cortex and parietal cortex in the form of 1- and 2-dimensional persistent cycles [31,46].

3.14   Music

Among other mathematical applications to music, M.G. Bergomi in Lisbon collaborates with various researchers in exploring musical genres by persistence [6]. As a space they adopt a modified version of Euler's *Tonnetz* [9]. The filtering function is the total duration of each note in a given track. Classification can be performed at different detail levels: experimentation is reported on tonal and atonal classical music of several authors (an example is in Fig. 15), on pop music and on different interpretation of the same jazz piece.

A blend of persistence and deep learning is the central idea of a research by the team of I.-H. Yang in Taiwan [38]. They input audio signals to a Convolutional Neural Network (CNN); after a first convolution layer, a middle layer
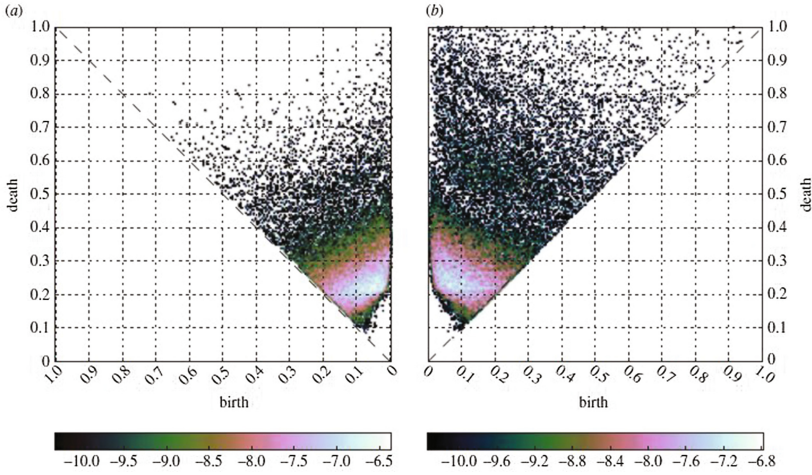
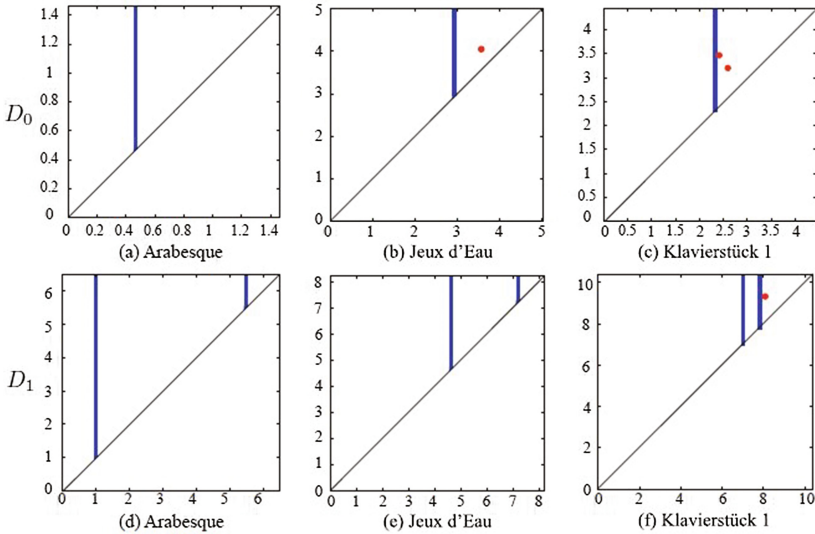**Fig. 14.** Probability densities for H_1 generators: placebo (left) and psilocybin (right) treated.

Fig. 15. 0- and 1-persistence diagrams for three classical pieces.

processes the output of the first in two different complementary ways: one is a classical CNN; the other computes the persistence landscape (an information piece derivable from the persistence diagram [10]) of the same output. Whereas the persistence layer by itself does not perform any better than the normal CNN, their combination gives very good results in terms of music tagging.

3.15 Languages

An interdisciplinary team at Caltech investigates the metric spaces built by 79 Indo-European and 49 Niger-Congo languages [44]. These appear as points in a Euclidean space of syntactic parameters; on them a Vietoris-Rips complex [23, Sect. III.2] is built and Euclidean distance is assumed as filtering function. The Indo-European family reveals one 1-dimensional and two 0-dimensonal persistent cycles, the Niger-Congo respectively none and one. The interpretation of these differences and of the link with phylogenetic and historical facts is still under way.

4   Open Problems

There is a number of open problems in persistence, whose solution will affect applications to natural data analysis, and to which only partial answers have been given so far:

– Optimal choice of the foliations along which to perform the 1D reduction of multidimensional persistence [13]
– Study of the discontinuities in multidimensional persistence [11,15]
– Understanding the monodromy around multiple cornerpoints [14]
– Restricting the group of homeomorphisms of interest by considering the invariance required by the observer [29]
– Modulation of the impact of different filtering functions for search engines with relevance feedback [30]
– Use of advanced tools of algebraic topology [5]
– Use of persistence in the wider context of concrete categories, not necessarily passing through homology of complexes or of topological spaces [7].

5   Future Outlook

There are at least two ways in which persistence will interact with machine learning, and this is likely to enormously boost the qualitative processing of natural data [18]:

– Feeding a neural network with Persistence Diagrams instead of raw data will convey the needs and viewpoints of the user
– Deep learning might yield a quantum leap in persistence, by automatically finding the best filtering functions for a given problem.

References

1. Adcock, A., Rubin, D., Carlsson, G.: Classification of hepatic lesions using the matching metric. Comput. Vis. Image Underst. **121**, 36–42 (2014)
2. Banerjee, S.: Size functions in the study of the evolution of cyclones. Int. J. Meteorol. **36**(358), 39 (2011)
3. Banerjee, S.: Size functions in galaxy morphology classification. Int. J. Comput. Appl. **100**(3), 1–4 (2014)
4. Bazaikin, Y.V., Baikov, V.A., Taimanov, I.A., Yakovlev, A.A.: Chislennyi analiz topologicheskih harakteristik trehmernyh geologicheskih modelei neftegazovyh mestorozhdenii. Matematicheskoe Modelirovanie **25**(10), 19–31 (2013)
5. Belchí, F., Murillo, A.: A$_\infty$-persistence. Appl. Algebra Eng. Commun. Comput. **26**(1–2), 121–139 (2015)
6. Bergomi, M.G., Baratè, A., Di Fabio, B.: Towards a topological fingerprint of music. In: Bac, A., Mari, J.I. (eds.) CTIC 2016. LNCS, vol. 9667, pp. 88–100. Springer, Cham (2016). doi:10.1007/978-3-319-39441-1_9
7. Bergomi, M.G., Ferri, M., Zuffi, L.: Graph persistence. arXiv preprint arXiv:1707.09670 (2017)
8. Biasotti, S., Cerri, A., Frosini, P., Giorgi, D., Landi, C.: Multidimensional size functions for shape comparison. J. Math. Imag. Vis. **32**(2), 161–179 (2008)
9. Bigo, L., Andreatta, M., Giavitto, J.-L., Michel, O., Spicher, A.: Computation and visualization of musical structures in chord-based simplicial complexes. In: Yust, J., Wild, J., Burgoyne, J.A. (eds.) MCM 2013. LNCS (LNAI), vol. 7937, pp. 38–51. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39357-0_3
10. Bubenik, P., Dłotko, P.: A persistence landscapes toolbox for topological statistics. J. Symb. Comput. **78**, 91–114 (2017)
11. Carlsson, G., Zomorodian, A.: The theory of multidimensional persistence. Discr. Comput. Geom. **42**(1), 71–93 (2009)
12. Carlsson, G., Zomorodian, A., Collins, A., Guibas, L.J.: Persistence barcodes for shapes. IJSM **11**(2), 149–187 (2005)
13. Cerri, A., Di Fabio, B., Ferri, M., Frosini, P., Landi, C.: Betti numbers in multidimensional persistent homology are stable functions. Math. Methods Appl. Sci. **36**(12), 1543–1557 (2013)
14. Cerri, A., Ethier, M., Frosini, P.: A study of monodromy in the computation of multidimensional persistence. In: Gonzalez-Diaz, R., Jimenez, M.-J., Medrano, B. (eds.) DGCI 2013. LNCS, vol. 7749, pp. 192–202. Springer, Heidelberg (2013). doi:10.1007/978-3-642-37067-0_17
15. Cerri, A., Frosini, P.: Necessary conditions for discontinuities of multidimensional persistent betti numbers. Math. Methods Appl. Sci. **38**(4), 617–629 (2015)
16. Chung, M.K., Bubenik, P., Kim, P.T.: Persistence diagrams of cortical surface data. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 386–397. Springer, Heidelberg (2009). doi:10.1007/978-3-642-02498-6_32
17. d'Amico, M., Ferri, M., Stanganelli, I.: Qualitative asymmetry measure for melanoma detection. In: IEEE International Symposium on Biomedical Imaging: Nano to Macro, pp. 1155–1158. IEEE (2004)
18. Dehmer, M., Emmert-Streib, F., Pickl, S., Holzinger, A.: Big Data of Complex Networks. CRC Press, Boca Raton (2016)
19. Donatini, P., Frosini, P.: Lower bounds for natural pseudodistances via size functions. Arch. Inequal. Appl. **1**(2), 1–12 (2004)

20. Donatini, P., Frosini, P.: Natural pseudodistances between closed manifolds. Forum Mathematicum **16**(5), 695–715 (2004)
21. Donatini, P., Frosini, P.: Natural pseudodistances between closed surfaces. J. Eur. Math. Soc. **9**(2), 231–253 (2007)
22. Edelsbrunner, H., Harer, J.: Persistent homology–a survey. In: Surveys on Discrete and Computational Geometry, vol. 453, pp. 257–282, Providence, RI (2008). Contemp. Math. Amer. Math. Soc
23. Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. American Mathematical Society, Providence (2009)
24. Ferri, M., Frosini, P., Lovato, A., Zambelli, C.: Point selection: a new comparison scheme for size functions (with an application to monogram recognition). In: Chin, R., Pong, T.-C. (eds.) ACCV 1998. LNCS, vol. 1351, pp. 329–337. Springer, Heidelberg (1997). doi:10.1007/3-540-63930-6_138
25. Ferri, M., Gallina, S., Porcellini, E., Serena, M.: On-line character and writer recognition by size functions and fuzzy logic. In: Proceedings of ACCV 1995, pp. 5–8 (1995)
26. Ferri, M., Lombardini, S., Pallotti, C.: Leukocyte classifications by size functions. In: Proceedings of the Second IEEE Workshop on Applications of Computer Vision, pp. 223–229. IEEE (1994)
27. Ferri, M., Stanganelli, I.: Size functions for the morphological analysis of melanocytic lesions. J. Biomed. Imaging **2010**, 5 (2010)
28. Ferri, M., Tomba, I., Visotti, A., Stanganelli, I.: A feasibility study for a persistent homology-based k-nearest neighbor search algorithm in melanoma detection. J. Math. Imaging Vis. **57**, 1–16 (2016)
29. Frosini, P., Jabłoński, G.: Combining persistent homology and invariance groups for shape comparison. Discrete Comput. Geom. **55**(2), 373–409 (2016)
30. Giorgi, D., Frosini, P., Spagnuolo, M., Falcidieno, B.: 3D relevance feedback via multilevel relevance judgements. Vis. Comput. **26**(10), 1321–1338 (2010)
31. Giusti, C., Pastalkova, E., Curto, C., Itskov, V.: Clique topology reveals intrinsic geometric structure in neural correlations. Proc. Natl. Acad. Sci. **112**(44), 13455–13460 (2015)
32. Handouyahia, M., Ziou, D., Wang, S.: Sign language recognition using moment-based size functions. In: Proceedings of International Conference on Vision, Interface, pp. 210–216 (1999)
33. Hatcher, A.: Algebraic Topology. Cambridge University Press, New York (2001)
34. Holzinger, A.: On knowledge discovery and interactive intelligent visualization of biomedical data. In: Proceedings of the International Conference on Data Technologies and Applications DATA 2012, Rome, Italy, pp. 5–16 (2012)
35. Holzinger, A.: On topological data mining. In: Holzinger, A., Jurisica, I. (eds.) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. LNCS, vol. 8401, pp. 331–356. Springer, Heidelberg (2014). doi:10.1007/978-3-662-43968-5_19
36. Kelly, D., McDonald, J., Lysaght, T., Markham, C.: Analysis of sign language gestures using size functions and principal component analysis. In: Machine Vision and Image Processing Conference, IMVIP 2008. International, pp. 31–36. IEEE (2008)
37. Lamar-León, J., García-Reyes, E.B., Gonzalez-Diaz, R.: Human gait identification using persistent homology. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) CIARP 2012. LNCS, vol. 7441, pp. 244–251. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33275-3_30

38. Liu, J.-Y., Jeng, S.-K., Yang, Y.-H.: Applying topological persistence in convolutional neural network for music audio signals. arXiv preprint arXiv:1608.07373 (2016)
39. Lord, L.-D., Expert, P., Fernandes, H.M., Petri, G., Van Hartevelt, T.J., Vaccarino, F., Deco, G., Turkheimer, F., Kringelbach, M.L.: Insights into brain architectures from the homological scaffolds of functional connectivity networks. Front. Syst. Neurosci. **10**, 85 (2016)
40. Micheletti, A.: The theory of size functions applied to problems of statistical shape analysis. In: S4G-International Conference in Stereology, Spatial Statistics and Stochastic Geometry, pp. 177–183. Union of Czech Mathematicians and Physicists (2006)
41. Micheletti, A., Landini, G.: Size functions applied to the statistical shape analysis and classification of tumor cells. In: Bonilla, L.L., Moscoso, M., Platero, G., Vega, J.M. (eds.) ECMI 2006. Mathematics in Industry, vol. 12, pp. 538–542. Springer, Heidelberg (2008)
42. Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P.J., Vaccarino, F.: Homological scaffolds of brain functional networks. J. Roy. Soc. Interface **11**(101), 20140873 (2014)
43. Platt, D.E., Basu, S., Zalloua, P.A., Parida, L.: Characterizing redescriptions using persistent homology to isolate genetic pathways contributing to pathogenesis. BMC Syst. Biol. **10**(1), S10 (2016)
44. Port, A., Gheorghita, I., Guth, D., Clark, J.M., Liang, C., Dasu, S., Marcolli, M.: Persistent topology of syntax. arXiv preprint arXiv:1507.05134 (2015)
45. Ringermacher, H.I., Mead, L.R.: A new formula describing the scaffold structure of spiral galaxies. Mon. Not. R. Astron. Soc. **397**(1), 164–171 (2009)
46. Sizemore, A., Giusti, C., Betzel, R.F., Bassett, D.S.: Closures and cavities in the human connectome. arXiv preprint arXiv:1608.03520 (2016)
47. Stanganelli, I., Brucale, A., Calori, L., Gori, R., Lovato, A., Magi, S., Kopf, B., Bacchilega, R., Rapisarda, V., Testori, A., Ascierto, P.A., Simeone, E., Ferri, M.: Computer-aided diagnosis of melanocytic lesions. Anticancer Res. **25**(6C), 4577–4582 (2005)
48. Turner, K., Mukherjee, S., Boyer, D.M.: Persistent homology transform for modeling shapes and surfaces. Inf. Inference J. IMA **3**(4), 310–344 (2014)
49. Uras, C., Verri, A.: On the recognition of the alphabet of the sign language through size functions. In: Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 2-Conference B: Computer Vision & Image Processing, vol. 2, pp. 334–338. IEEE (1994)
50. Verri, A., Uras, C., Frosini, P., Ferri, M.: On the use of size functions for shape analysis. Biol. Cybern. **70**, 99–107 (1993)
51. Ziefle, M., Himmel, S., Holzinger, A.: How usage context shapes evaluation and adoption criteria in different technologies. In: AHFE 2012, Proceeding of International Conference on Applied Human Factors and Ergonomics, San Francisco, pp. 2812–2821 (2012)