Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning

(Article begins on next page)

26 April 2024

Towards a topological-geometrical theory of group equivariant non-expansive operators for data analysis and machine learning

Mattia G. Bergomi[a], Patrizio Frosini[b,c,*], Daniela Giorgi[d], Nicola Quercioli[b,c]

[a]*Champalimaud Research, Champalimaud Center for the Unknown - Lisbon, Portugal*
[b]*Department of Mathematics, University of Bologna*
[c]*Advanced Research Center on Electronic System "Ercole De Castro", University of Bologna*
[d]*Italian National Research Council, Institute of Information Science and Technologies "Alessandro Faedo"*

Abstract

We provide a general mathematical framework for group and set equivariance in machine learning. We define group equivariant non-expansive operators (GENEOs) as maps between function spaces associated with groups of transformations. We study the topological and metric properties of the space of GENEOs to evaluate their approximating power and set the basis for general strategies to initialise and compose operators. We define suitable pseudo-metrics for the function spaces, the equivariance groups, and the set of non-expansive operators. We prove that, under suitable assumptions, the space of GENEOs is compact and convex. These results provide fundamental guarantees in a machine learning perspective. By considering isometry-equivariant non-expansive operators, we describe a simple strategy to select and sample operators. Thereafter, we show how selected and sampled operators can be used both to perform classical metric learning and to inject knowledge in artificial neural networks.

Keywords: Group equivariant non-expansive operator, invariance group, natural pseudo-distance, group action, persistent homology, persistence diagram, bottleneck distance, agent, perception pair, topological data analysis
2000 MSC: Primary 55N35 Secondary 47H09, 54H15, 57S10, 68U05, 65D18

Introduction

Deep learning-based algorithms have reached human or superhuman performance in many real-world tasks (see, e.g. [1, 2]). Currently, Convolutional Neural Networks (CNNs) [3] are one of the most successful and flexible deep learning architectures, used in a range of applications spanning from image [4] to sentence classification [5], among others. CNNs are

---

*Corresponding author
Email addresses:* `mattia.bergomi@neuro.fchampalimaud.org` (Mattia G. Bergomi),
`patrizio.frosini@unibo.it` (Patrizio Frosini), `daniela.giorgi@isti.cnr.it` (Daniela Giorgi),
`nicola.quercioli2@unibo.it` (Nicola Quercioli)

characterised by the election of convolution as the operator of choice to act on the data. The convolutional kernels learned by optimising a loss function are operators that map the input data to new representations that are, for instance, more easily classifiable.

We provide a formal mathematical framework, in which data are represented as function spaces endowed with characteristics that allow us to guarantee basic results for learning, e.g. approximability and convexity. We do this by studying geometrical and topological properties of a space of operators acting on the function space of data, that are naturally robust to geometrical transformations (e.g. rotation), and reduce the dimensionality of the input data. While we acknowledge the relevance of statistical properties of data for machine learning [6], such as its distribution, we do not consider this here.

Equivariance in machine learning. An operator is called equivariant with respect to a group if the action of the group commutes with the operator. One of the main strengths of CNNs is the natural equivariance of the convolution operator with respect to interpretable transformations, namely the group of planar translations, at least in the ideal, continuous case. However, when working with images, volumes or even time series, oftentimes invariance with respect to transformations such as rotations, reflexions, or other deformations is fundamental to speed up the learning process, or even to reach satisfying accuracy. Currently, data augmentation, or heavy preprocessing (e.g. accurate image alignment) are the most common strategies used to produce networks resistant to even simple data transformations. Recently, there has been a growing interest for invariant representations in machine learning [7, 8, 9]. The reason is twofold. On one hand, the use of operators equivariant with respect to specific transformations allows one to inject pre-existing knowledge in the system, thus gaining control of the nature of the learned operators [10]. On the other hand, equivariance with respect to the action of a group (or a set) of transformations corresponds to the introduction of symmetries in the data space, hence drastically reducing the dimensionality of the space to be explored during optimisation, and opening the way to alternative kinds of abstract representation. Indeed, it is well known that incorporating prior domain knowledge helps machine learning.

Ours is not the first attempt to describe the role of mathematical transformations in the context of machine learning. From an information-theoretic viewpoint, the goal of machine learning can be described as a tradeoff between compression (i.e. the generation of meaningful, low-dimensional data representations) and prediction [11]. In [12, 13], invariant and non-expansive operators are studied to reduce the data variability in directions of local symmetries, using the language of filters. Finally, equivariant transformations in deep learning are studied in [14, 7, 15], by focusing on learning symmetries.

Our contribution is the definition of a mathematical model where data are represented as function spaces associated with groups of transformations. Data are manipulated through Group Equivariant Non-Expansive Operators (GENEOs) [16], which are in layman's terms, blind to the action of the group on the data. We prove that, under the assumption that the function spaces are compact, the space of GENEOs is compact with respect to a suitable pseudo-metric. The compactness guarantees that any operator belonging to a certain space can be approximated by a finite number of operators sampled in the same space. Also,

we prove that, under the assumption that the function spaces are convex, the spaces of GENEOs are convex. The convexity guarantees that we can generate new operators by convex combination of existing ones. Finally, we define and discretise a family of isometry equivariant non-expansive operators (IENEOs), that will be used in the presented applications on image classification.

We believe our research complements the current approaches. Our framework rests on a clear epistemological approach, focusing on the analysis of the data-agent pairs, rather than the data *per se*. Agents are GENEOs operating on data represented as samples of a function space. Thus, data are visible to a GENEO only under its action, which generates stable, compressed representations of data, while respecting the chosen equivariance constraint. Moreover, our framework applies to any kind of data that can be described by real or vector valued functions—including images, 3D data, and time series—and is grounded on the topological and geometrical study of the space of *all* group-equivariant non-expansive operators. Moving from the study of single operators to the metric analysis of the space of all GENEOs opens the way for discovering general principles in machine learning, instead of focusing on particular methods. The formal proofs of the compactness, approximability and convexity of the space of GENEOs under suitable assumptions are examples of results produced in this general setting. Consider for instance how requesting non-expansivity of operators translates into the quest for meaningful compressions of the input data. In the framework we propose, this request becomes a necessary hypothesis to build compact spaces of operators, where one can get effective representations of the input data through a finite number of operators.

Furthermore, the topological and metric properties we prove in the space of GENEOs set the basis for general strategies to initialise and compose operators. In this respect, it is important to notice how our framework allows for the definition of operations between operators, whose function spaces and groups of equivariance are different. This feature is fundamental, because it allows one to compose operators to form networks, in the same fashion as computational units are connected in an artificial neural network. Finally, an important advantage of having a general mathematical theory of GENEOs is the availability of simple and practical methods to build group equivariant non-expansive operators [17], without integrating on large groups [18].

Topological data analysis and machine learning. Two main challenges need to be tackled to implement our framework in a machine learning context. First, the space of operators has to be endowed with an easily computable metric, to be *rapidly* explored. Moreover, data under the action of operators must be comparable *via* an efficient metric. Indeed, it is necessary to quantify and compare the performance of operators, to learn and choose the ones compatible with a given task. While several methods to compare data represented by functions are illustrated in literature (see, e.g., the concept of derived kernel [19]), we propose the use of existing and novel tools in topological data analysis (TDA), and in particular persistent homology, to define computationally efficient metrics for comparing both GENEOs and transformed data. We do this by taking advantage of the desirable computational properties of TDA, deriving from its efficiency in extracting relevant information from data.

TDA is an emerging field of research which studies topological approaches to explore and provide meaningful, tight representations of complex, high-dimensional data drawn for instance from artificial and biological networks [20, 21]. The basic idea is that topology can help to recognise patterns within data, and therefore to turn data into compressed, useful knowledge. One of the main concepts in TDA is persistent homology (PH), a mathematical tool that captures topological information at multiple scales, and enables a fast comparison of functions defined on data.

The synergy between TDA and machine learning is currently studied along two main research lines. On one hand, topological and geometrical features of the data are extracted *via* TDA before applying classical machine learning techniques [22, 23]. On the other hand, recent works have shown how the analysis of the topology of the space of filters learned by CNNs can give insights into the internal representation of knowledge acquired by different networks [24, 25]. Our approach opens a new research direction. Indeed, we use tools in TDA to define *novel* metrics to navigate the space of operators. Instead of studying the topology of the internal representation of a given CNN, we consider the entire space of group equivariant non-expansive operators, endowed with a suitable pseudo-metric, and study its topological and geometrical properties from a theoretical point of view.

In particular, we propose an integration between the theory of group actions and persistent homology to define a strongly group-invariant pseudo-metric to compare data under the action of operators. We also define a pseudo-metric between GENEOs. The two pseudo-metrics are used to define two algorithms providing strategies to select and sample from a space of operators, given a dataset labelled for a classification task. These procedures allow one to first select a subset of operators belonging to a certain GENEOs space, which give a meaningful, compressed representation of data with respect to their labelling, always invariant under the transformations induced by the action of the invariance group. Thenceforth, the sampling algorithm allows one to eliminate redundant operators.

Applications. We show how the selection and sampling strategy can be used to perform metric learning on MNIST and fashion-MNIST. In addition, we show how convolutional filters initialised by selecting and sampling on few samples effectively grasp useful knowledge, which can be utilised to classify the remainder of the samples, for instance by a dense classifier.

A major advantage brought by the introduction of persistent homology is that the approach can be generalised to obtain equivariance with respect to a set of transformations, instead of to a group of transformations. This property is fundamental for those applications which require operators equivariant with respect to a finite set of transformations that is not closed with respect to composition and computation of the inverse, for example, isometric deformations only up to a certain degree.

Results

Theoretical results

Our mathematical approach is based on the choice of the sup-norm metric D_Φ on the function space Φ. According to the central role of measurements in our epistemological

setting, we introduce the pseudo-metric $D_X(x_1, x_2) = \sup_{\varphi \in \Phi} |\varphi(x_1) - \varphi(x_2)|$ on the set X and the pseudo-metric $D_G(g_1, g_2) := \sup_{\varphi \in \Phi} D_\Phi(\varphi \circ g_1, \varphi \circ g_2)$ on the group G, so that all the topologies and metrics used in our model are inherited by the topology and metric used to compare the data we are interested in. We recall that a pseudo-metric is just a distance d without the property that $d(a, b) = 0$ implies $a = b$. The group G acts on Φ by right composition and we can prove that this action is continuous (cf. **Supplementary Methods**, Theorem 2.7).

Our model is based on the study of the topological and geometrical properties of the space of GENEOs, seen as mathematical descriptions of agents acting on data. Each GENEO F is an operator that *a)* transforms the data represented by the functions belonging to a space Φ into functions belonging to another, possibly different, space Ψ, and *b)* transforms the group G acting on Φ into a group H acting on Ψ. After fixing a homomorphism $T : G \to H$, we require that the operator F is equivariant, i.e. $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for every $\varphi \in \Phi, g \in G$, and non-expansive, i.e. $D_\Psi(F(\varphi_1), F(\varphi_2)) \leq D_\Phi(\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$.

We can endow the set of all GENEOs with the pseudo-metric $D_{\mathrm{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} D_\Psi(F_1(\varphi), F_2(\varphi))$. Our main theoretical results consist in proving that, under the assumption that the function spaces are compact and convex, the topological spaces of all GENEOs are compact and convex too (Theorems 7 and 8). It is important to observe that, because of its compactness, every space of GENEOs can be approximated with arbitrary precision by a finite set of operators.

We are interested in comparing the functions in Φ as seen by a set $\mathcal{F}$ of GENEOs, i.e. in studying the pseudo-metric $D_{\mathcal{F},\Phi}(\varphi_1, \varphi_2) := \sup_{F \in \mathcal{F}} D_\Psi(F(\varphi_1), F(\varphi_2))$, for $\varphi_1, \varphi_2 \in \Phi$. The computation of the pseudo-metrics $D_{\mathcal{F},\Phi}$ and D_{GENEO} can be expensive, but persistent homology allows us to replace such pseudo-distances with two suitable pseudo-metrics $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ and Δ_{GENEO}. The definitions of $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ and Δ_{GENEO} are given by linking the theory of GENEOs to TDA, by means of the comparison of persistence diagrams via the matching distance δ_{match}. For details on the matching distances see **Supplementary Methods, Background on persistent homology**. The metrics $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ and Δ_{GENEO} are defined in Methods, Eq. 9 and 12, respectively. As a consequence, the efficient and powerful machinery developed in persistent homology can be used in our mathematical approach as a proxy for the comparison of functions and the metric analysis of the spaces of GENEOs. We observe that $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$ is stable and strongly invariant with respect to the action of the group G (Proposition 9).

Further properties of the pseudo-metrics and topologies used in this paper can be found in the **Supplementary Methods**.

Applications and experiments

Our mathematical model and theorems are based on the assumption that data can be treated as points in a space of continuous functions. We test the validity of such results on the classification of real-world datasets, namely, the image datasets MNIST [28], fashion-MNIST [29] and CIFAR-10 [30]. We first define a parametric family of discrete non-expansive operators which are equivariant with respect to Euclidean plane isometries. Then, we define an algorithm enabling the selection and sampling of GENEOs in order to learn the metric
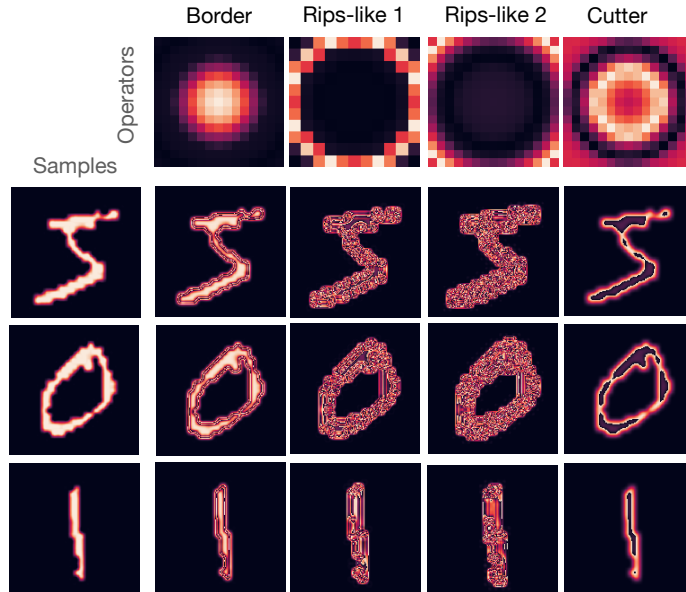
Figure 1: **Isometry equivariant non-expansive operators on MNIST.** Although the operator space is constrained by the symmetries induced by the equivariance with respect to the group of planar isometries, we observe that the operators are expressive and have a clear topological interpretation. Among other functions, these example operators are specialised in border detection and voids (cycles in topological terms) filling, similar to the classical Vietoris-Rips construction [26]. Finally, the cutter operator–depicted in the last column–disconnects the image approximately according to its distance transform [27].

induced on a dataset by a labelling function. We show that selection and sampling can approximate an agent able to express the underlying metric of the datasets by observing only 20 or 40 examples per class. Thereafter, we show how the metric learned is still expressive when used to represent distances among validation samples transformed according to the equivariances of the GENEOs of choice. Finally, we use selected and sampled GENEOs to inject knowledge in an artificial neural network.

Isometry Equivariant Non-Expansive Operators (IENEOs). We define and discretise a parametric family of non-expansive operators which are equivariant with respect to Euclidean plane isometries. We set $(\Phi, I) = (\Psi, I)$, and $\Phi = \{\varphi : \mathbb{R}^2 \to \mathbb{R}, \text{ continuous, with compact support}\}$ and T the identity homomorphism. Given a continuous, real-valued, non-null function $g : \mathbb{R} \to \mathbb{R}$ with compact support, we get a function $G : \mathbb{R}^2 \to \mathbb{R}$ by taking $G(x, y) = g\left(\sqrt{x^2 + y^2}\right)$, normalized by its L^1 norm to guarantee non-expansivity. Then, the convolutional operator F mapping each continuous function with compact support $\varphi : \mathbb{R}^2 \to \mathbb{R}$ to $F(\varphi) = \varphi * G$ is a group equivariant non-expansive operator with respect to the group I of Euclidean plane isometries. We generate a parametric family of IENEOs by considering g as a convex combination of Gaussians parametrised by their widths and centres. See Methods, Section Isometry equivariant non-expansive operators, for the formal definition of IENEOs. Figure 1 shows some IENEOs, discretised as kernels and applied to images in the MNIST dataset. The examples include operators with a clear semantic interpretation, always equivariant with
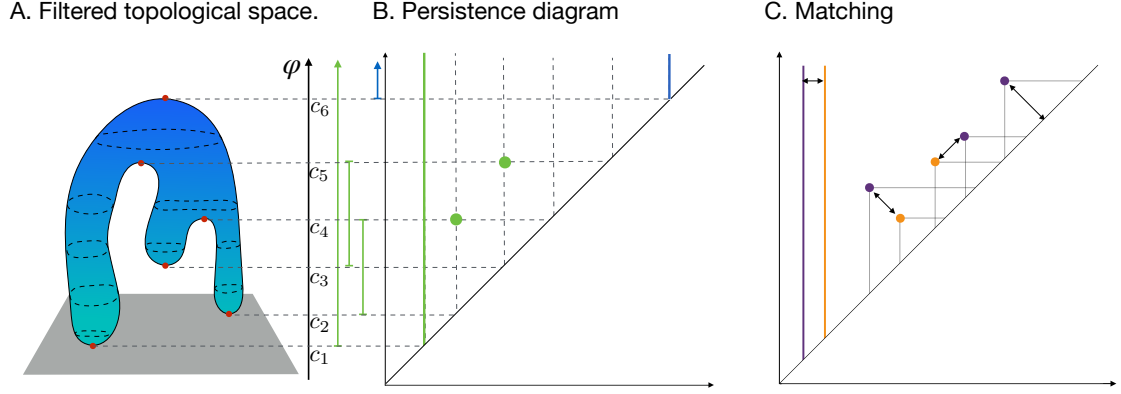
Figure 2: **Persistent homology background.** In persistent homology we consider pairs composed by a topological space and a continuous function defined on the topological space of interest. The (homological) critical values of the function induce naturally a sublevel set filtration of the topological space. In panel A, a topological sphere is filtered by considering the critical values of the height function. We obtain a filtration by considering the sequence of nested sublevel sets ordered according to the natural order on the critical values. The evolution throughout the filtration of the number of generators of the kth homology groups (i.e. the number of k-dimensional holes) is represented as a persistence diagram in Panel B. 0-dimensional holes, or connected components are represented as green cornerpoints. The void generated when considering the last sublevel set, corresponding to the entire space, generates the cornerline depicted in blue. A distance between two persistence diagrams can be computed as an optimal matching of cornerpoint. The matching process is depicted in Panel C. Note how non-matchable cornerpoints can be associated to their projection on the diagonal.

respect to rotation, translation and reflexion.

Operators selection and sampling on labelled datasets. We start from the assumption that data with the same label share common features with respect to the agent we want to approximate. Thus, we suggest an algorithm for metric learning based on the metrics $\mathcal{D}^{\mathcal{F},k}_{\text{match}}$ and Δ_{GENEO} we introduced on the space of GENEOs grounded on persistent homology, see Methods, Equations (9) and (12).

Briefly, we start by selecting randomly a certain number of operators, IENEOs in our experiments. Afterwards, we apply them to the labelled images, and compare the transformed images using the matching distance between their persistence diagrams. In this way, we take advantage of the fact that the representation of images as persistence diagrams (in Figure 2) is invariant with respect to the action of the equivariance group of the operators, planar isometries in our case. Then, we select operators that have small distances for the samples associated to the same label.

The selection criterion does not guarantee that the operators are maximally diverse, when evaluated within and between classes. The important advantage of working on metric spaces is that we can sample the selected elements to avoid storing operators that would focus on the same or similar feature. Always profiting from the property of the matching distance to be lower bound of the metric defined on the space of operators, we sample the operators using the metric Δ_{GENEO}. Thereafter, by sampling operators which are distant from each other, we obtain a minimal set of non-redundant operators. The selection and sampling
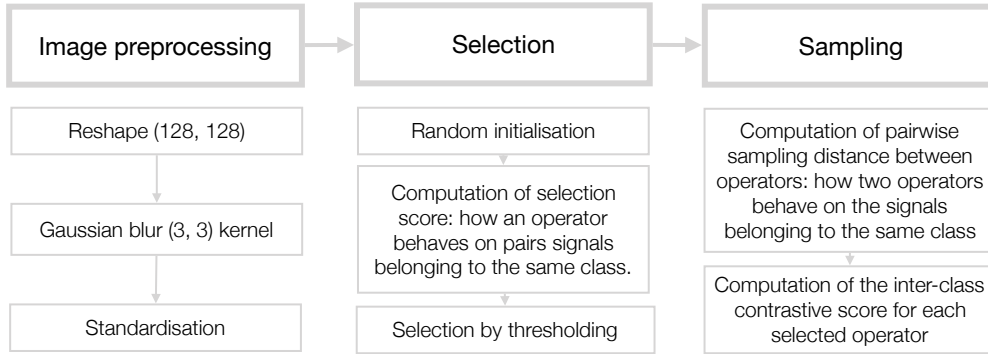
Figure 3: Experimental pipeline. In the preprocessing step images are reshaped, smoothed and standardised. Subsequently, operators are initialised randomly. Thereafter, they are selected according to their output when evaluated on objects belonging to a chosen class. The final sampling step allows us to exclude operators that appear to be too similar, and thus redundant.

algorithm is formally described in Methods, Sec. Experimental Setting.

Operators that have been selected and sampled on labelled examples are expected to grasp geometrical and topological features characterising samples belonging to the same class. Thus, metric learning is a natural application, as shown below.

Metric learning. We use selected and sampled operators to measure distances between pairs of validation samples, *via* the metric $\mathcal{D}_{\mathrm{match}}^{\mathcal{F},k}$. This choice implies that two samples have distance 0, and hence are considered the same by the collection of selected operators (agents), only if every operator *sees* them as identical. Again, the invariance with respect to the action of the group of planar isometries is naturally inherited by the usage of metrics based on persistent homology.

After computing the pairwise distance induced by the selected and sampled operators between validation samples, we use hierarchical clustering [32] to visualise in a dendrogram, how samples have been organised by the learned metric. We find that few examples per class are enough to learn a metric that generalises to new validation samples, and solve a binary classification task on the MNIST dataset.

We proceed as follows:

1. Select a dataset $D \in \{\mathrm{MNIST}, \mathrm{fashion\text{-}MNIST}, \mathrm{CIFAR\text{-}10}\}$
2. Select two random classes l_1, l_2 among the classes labelling D and consider the sets Φ_1, Φ_2 of samples associated to the selected classes.
3. Preprocess the images according to the procedure described in Figure 3
4. Choose randomly $n << |\Phi_i|$ samples per class ($|\Phi_i| = 6000$ for every i and $n \in \{20, 40\}$ in our experiments).
5. Initialise, select and sample IENEOs by evaluating them on the random samples extracted at the previous step.
6. Use the selected and sampled IENEOs to compute distances between validation samples (ten per class).

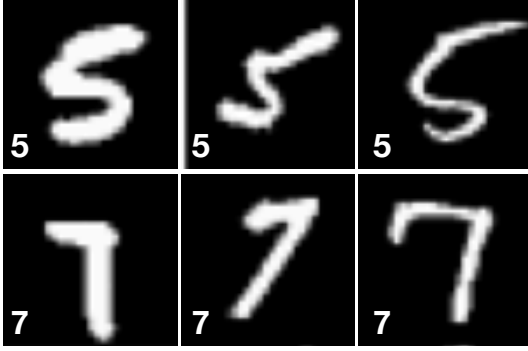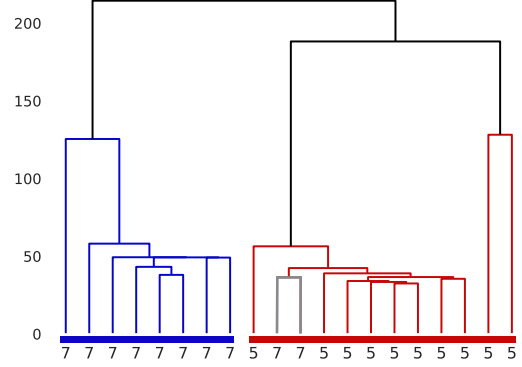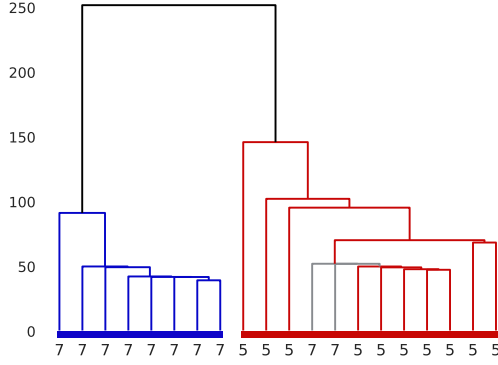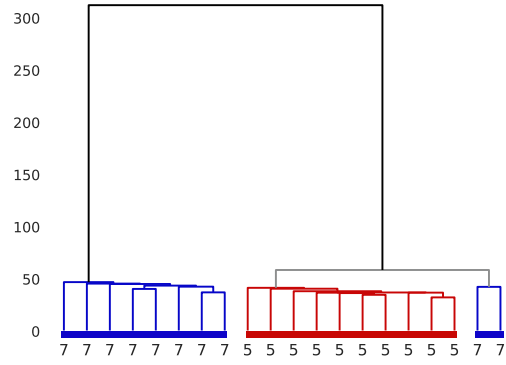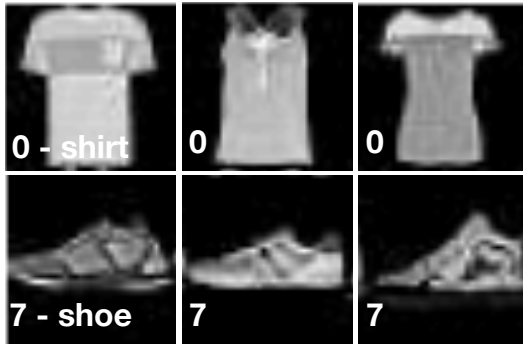Figure 4: **IENEO-based metric learning on MNIST and CIFAR-10.** A. Samples from the classes 7 and 5 from the MNIST dataset. Panels B, C and D show the hierarchical clustering obtained by using selected and sampled IENEOs of different dimensions to measure the distance between validation samples belonging to the two considered classes. For all filter sizes 500 operators were randomly initialised, then selected and sampled. We observe how samples belonging to the two classes are clearly separated by filters of all the considered dimensions.

We also test the impact of changes of the kernel size $s \in \{7\text{px}, 11\text{px}, 21\text{px}\}$, on the learned metric. Results on the three datasets, with different kernel sizes are displayed in Figures 4 and 5.

Generalisation to transformed samples. As an additional experiment, we evaluated the capacity of the selected and sampled operators to discriminate validation samples that have been transformed with random planar isometries. The validation on augmented sampling application aims at testing the aforementioned equivariance of the distance $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$. To do this, we consider a set of operators selected and sampled on non-transformed samples of MNIST, while we transform the set of validation samples by applying a random transformation among translations, rotations and reflexions parametrised as follows:

1. rotations are selected randomly to be between 1 and 30 degrees;
2. translations can be in both the x and y-axis directions in a range between 1 and 2 pixels;
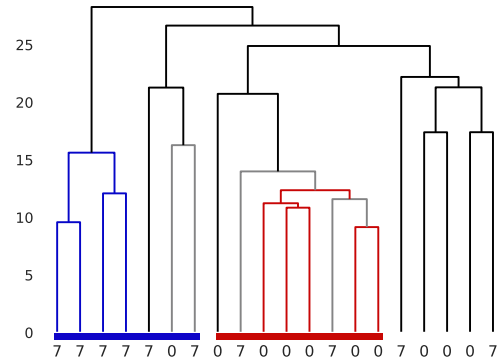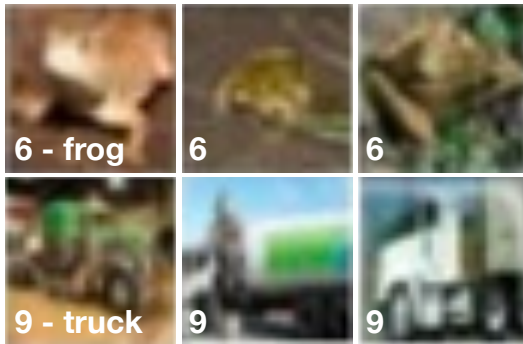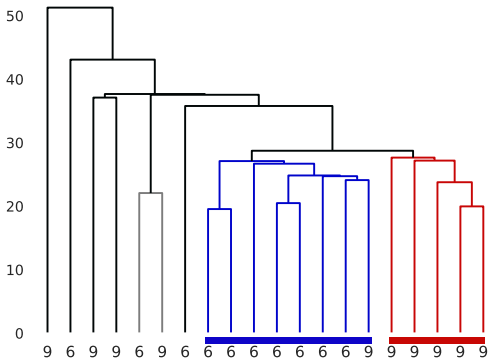
Figure 5: **IENEO-based metric learning on fashion-MNIST and CIFAR-10.** Panels A and C show some samples from the fashion-MNIST [29] and CIFAR-10 [31, 30] datasets. The dendrograms in Panels B and D represent as dendrograms the metric learnt after selecting and sampling 500 randomly initialised IENEOs on 20 examples per class. We evaluated the metric encoded by the selected operators on a validation set consisting of 10 examples per class.

3. reflexions are computed randomly with respect to one of the two axes.

The transformed samples along with the dendrograms obtained by considering the metric induced by the selected and sampled operators are shown in Figure 6.

Knowledge injection. As a final application, we discuss the possibility of using selected and sampled operators as feature extractors for a simple artificial neural network model. The aim of this experiment is twofold. First, we provide additional evidence that the filters obtained through the selection and sampling procedures are informative with respect to a chosen dataset, even when evaluated on few samples per class. Second, we aim at showing that this information can be used for classification and hence *injected* as prior knowledge in an artificial neural network architecture under the form of fixed convolutional kernels. The architecture of the model is shown in Figure 7.

After preprocessing the samples of MNIST, and applying the selection and sampling algorithms as described in the previous experiment, we consider as feature extractors three different sets of IENEOs generated by considering as many sets of initialisation parameters
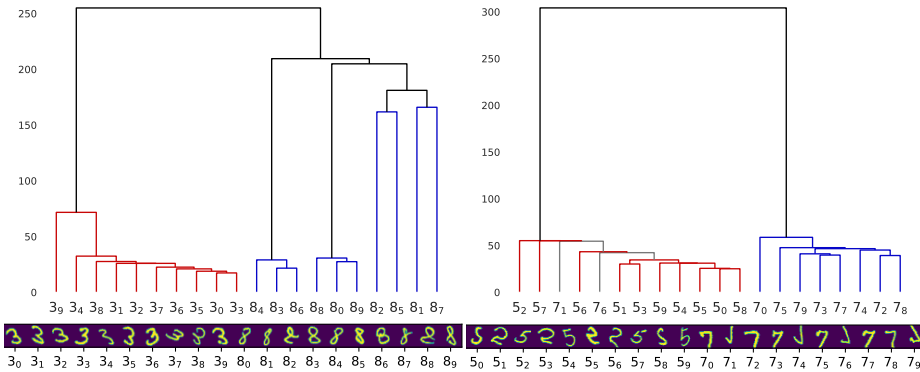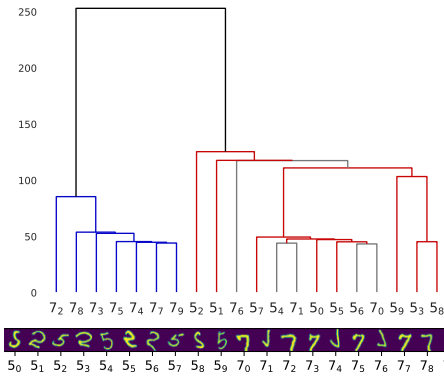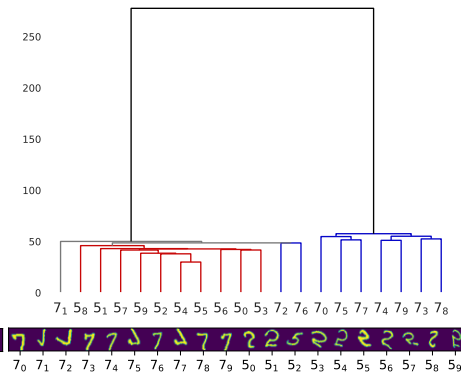
Figure 6: **Metric learning on transformed MNIST validation samples.** The metric obtained by considering selected and sampled IENEOs can be used to cluster samples transformed according to the group of equivariances (in this case planar isometries) of the family of operators of choice. In panel A we compare the cluster of transformed validation samples obtained on two different pairs of MNIST classes. The dendrogram on the right of panel A, panels B and C show how a variation in the size and number of Gaussian components of the IENEO affects the clustering of validation samples randomly transformed through planar isometries.

for selection and sampling. Each experiment is repeated three times per dataset and new classes are extracted randomly at each repetition. As a control, we compare the classification performance (accuracy on the validation set) of the classifiers with IENEO injected knowledge, with the same classifiers receiving features extracted with random filters [33].

Selected and sampled operators effectively carry relevant information about the dataset samples, as shown by the classification performance summarised in Table 1.

Discussion

The rationale of our approach is that the main interest in machine learning does not consist in the analysis and the approximation of data, but rather in the analysis and the approximation of the observers looking at the data. A simple example can make this idea clearer: if we consider images representing skin lesions, we are not mainly interested in the
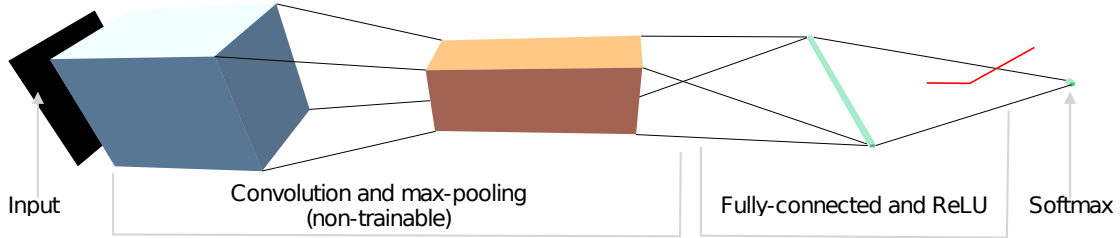
Figure 7: **Convolutional neural network architecture used in the knowledge injection experiment.** The two first layers are a convolutional and maxpooling layer, respectively. The parameters of these layers are fixed. Two fully-connected layers counting 64 and 2 units respectively are trained to classify the two selected classes. The first fully-connected layer uses a rectified linear unit (ReLU) and the second a softmax as activation functions.

images *per se*, but rather in approximating the judgement given by the physicians about such images.

With this motivation in mind, the first contribution of this paper consists in describing a formal framework for machine learning, based on the study of metric and topological properties of operator spaces acting on function spaces. Of all possible types of operators, we study the space of non-expansive, group equivariant operators (GENEOs). Being non-expansive, operators of this kind enable us to find compressed representations of the original input. Group equivariance, i.e. commutativity of the operator with respect to the action of a chosen group, makes these operators able to provide stable representations of the data, even when transformed by the action of elements of the groups.

Suitable pseudo-metrics allow us to define a topology on the function space of data and then induce a topology on the space of GENEOs. We build the necessary machinery to define operations between GENEOs whose groups of equivariance are different. This definition is fundamental, because it allows one to compose operators hierarchically, in the same fashion as computational units are linked in an artificial neural network. Thereafter, by taking advantage of known and novel results in persistent homology, we prove compactness and convexity of the space of GENEOs under suitable hypotheses. On the side of applications, we give two algorithms to select and sample from a space of GENEOs, given a dataset labelled for a classification task. These procedures allow one to first select a subset of operators that give meaningful representations of the data with respect to their labelling. Thenceforth, the sampling algorithm allows one to eliminate redundant operators. These two strategies are used to perform metric learning on the MNIST, fashion-MNIST and CIFAR-10 datasets. In addition, we show how convolutional filters initialised by selecting and sampling on few samples, effectively grasp useful knowledge that can be utilised to train and validate an artificial neural network.

We believe that our model, at the interface of functional analysis and topology, opens

Dataset	500 operators 20 examples	500 operators 40 examples	750 operators 20 examples	Random filters
MNIST	0.988591	0.963282	**0.991158**	0.506815
MNIST	**0.989732**	0.949096	0.987450	0.960501
MNIST	**0.990588**	0.960501	0.989447	0.506815
f-MNIST	0.984598	0.960223	**0.986309**	0.506815
f-MNIST	0.990017	0.958275	**0.990588**	0.958554
f-MNIST	**0.990302**	0.963561	0.988591	0.506815
CIFAR-10	0.982886	0.960223	**0.987735**	0.506815
CIFAR-10	**0.992299**	0.962170	0.988876	0.506815
CIFAR-10	0.982316	0.960501	**0.989732**	0.971627
MNIST $\mu(\sigma)$	0.989(0.001)	0.958(0.007)	0.989(0.001)	0.658(0.262)
f-MNIST $\mu(\sigma)$	0.988(0.003)	0.961(0.002)	0.988(0.002)	0.657(0.261)
CIFAR-10 $\mu(\sigma)$	0.985(0.005)	0.961(0.001)	0.988(0.001)	0.661(0.268)
All $\mu(\sigma)$	0.988(0.003)	0.959(0.004)	0.988(0.001)	0.659(0.228)

Table 1: **Evaluating the knowledge injection capability of GENEOs.** Comparison between the performance obtained by classifying two classes of MNIST, fashion-MNIST and CIFAR-10 with a dense classifier fed with features extracted by convolutional filters obtained by selecting and sampling IENEOs. Column headers specify the dataset and the initialisation parameters used to select and sample the IENEOs. The last column is the control for our experiments, where the classification is based upon features extracted with random filters. Experiments have been repeated three times for each dataset and initialisation parameters. In the columns, the accuracy obtained by the classifier on the validation set (3000 images per class) is reported, along with the dataset-wise and overall mean accuracy and standard deviation.

new avenues for further research. Our main forward-looking goal is the definition of a novel artificial neural network model based on functional modules. Modules would be more complex computational units than standard artificial neurons, with the same flavour of [34, 35]. The core of each module would be a collection of GENEOs, thus each module would be defined a priori to be equivariant with respect to a set of transformations. On one hand, this choice would result in a significant reduction of the dimensionality of the manifold to be studied during optimisation. On the other hand, choosing the transformation equivariance to be respected at each layer would allow us to inject knowledge in the networks before training, and would assure that information is not acquired by relying on unwanted noisy regularities in the training data. We recall that in this context, it will be necessary to investigate the minimal statistical properties that a dataset needs to obey in our framework. Module networks would learn optimal transformations of the data to achieve a task, rather than operating on data themselves. To achieve the goal, new methods for building GENEOs should be developed, in order to get good approximations of the spaces of GENEOs for given equivariance groups and function spaces. Moreover, as some sets of GENEOs appear to have a structure of a Lie group and a Riemannian manifold, these structures seem worthy of study

and analysis.

Acknowledgements

Author contributions

P.F., M.G.B., D.G., N.Q. devised the project and algorithms. P.F. and N.Q. formalised the mathematical model with help from M.G.B. and D.G.. M.G.B. wrote the code with help from D.G.. M.G.B. and D.G. analysed data. P.F. supervised the project. M.G.B., P.F., D.G., N.Q. contributed equally to the writing of main text and methods.

Software availability

Experiments and applications can be reproduced by installing the open-source Python package available at doi.org/10.5281/zenodo.3264851.

Data availability

No datasets were generated during the current study. The datasets analysed are available at

- MNIST [28], yann.lecun.com/exdb/mnist/;

- Fashion-MNIST [29], github.com/zalandoresearch/fashion-mnist;

- CIFAR-10 [30], www.cs.toronto.edu/ kriz/cifar.html.

Methods

Mathematical model

The aim of this section is to describe in a streamlined and intuitive manner the mathematical results detailed and proven in **Supplementary Methods, Mathematical model**.

Epistemological setting. Our mathematical model is justified by an epistemological background which revolves around the following assumptions:

1. Data are represented as functions defined on topological spaces, since only data that are stable with respect to a certain criterion (e.g., with respect to some kind of measurement) can be considered for applications, and stability requires a topological structure.
2. Data cannot be studied in a direct and absolute way. They are only knowable through acts of transformation made by an agent. From the point of view of data analysis, only the pair (data, agent) matters. In general terms, agents are not necessarily endowed with purposes or goals: they are just ways and methods to transform data. Acts of measurement are a particular class of acts of transformation, that can or not be at the service of a global goal quantified by a loss or a reward function.
3. Agents are described by the way they transform data while respecting some kind of invariance. In other words, any agent can be seen as a group equivariant operator acting on a function space.
4. Data similarity depends on the output of the considered agent.

In other words, in our framework we assume that the analysis of data is replaced by the analysis of the pair (data, agent). Since an agent can be seen as a group equivariant operator, from the mathematical viewpoint our purpose consists in presenting a good topological theory of suitable operators of this kind, representing agents. For more details, we refer the interested reader to [36].

Data representation. In our mathematical model, data are represented as function spaces, that is, as sets Φ of bounded real-valued functions $\varphi : X \to \mathbb{R}$, with X a set. We assume that Φ is compact with respect to the topology induced by the distance

$$D_\Phi(\varphi_1, \varphi_2) := \|\varphi_1 - \varphi_2\|_\infty . \tag{1}$$

We can think of X as the space where one makes measurements, and of Φ as the set of admissible measurements. For example, an image can be represented as a function φ from the real plane X to the real numbers. X inherits the structure of a topological space from the pseudo-metric D_X which distinguishes points only if they are seen as different by some measurement:

$$D_X(x_1, x_2) = \sup_{\varphi \in \Phi} |\varphi(x_1) - \varphi(x_2)| \tag{2}$$

for every $x_1, x_2 \in X$. We recall that a pseudo-metric is just a distance d without the property that $d(a, b) = 0$ implies $a = b$. The reason to consider a topological space X rather than considering just a set follows from the need of formalising the assumption that data are stable. Note that our choice of topology allows to deal with non-continuous functions, with respect to the Euclidean topology (see **Supplementary Methods**, Theorem 2.1).

In general X is not compact with respect to the topology induced by D_X, even if Φ is compact. For example, if X is the open interval $]0, 1[$ and Φ contains only the identity from $]0, 1[$ to $]0, 1[$, the topology induced by D_X is simply the Euclidean topology and hence X is not compact. However, the next result holds.

Theorem 1. *If Φ is compact and X is complete then X is also compact.*

Proof. Supplementary Methods, Theorem 2.2. □

In our paper we will assume that X is complete and hence compact.

Most types of data can be represented according with the hypotheses of our framework. Time series can be seen as functions defined on the real line. 3D objects represented as points clouds can be encoded as real functions defined on $\mathbb{R}^3$, for example via Radial Basis Functions approximation [37, 38]. Oftentimes, in data analysis, samples are represented as filtrations in metric spaces, i.e. a sample is seen as a sequence of nested sublevels with respect to a given property. See e.g. Panel A of Figure 2 and Vietoris-Rips filtration [39]. Our approach can be used without loss of generality in this case as well, since it has been shown that every compact and stable 1-dimensional filtration of a compact metric space is induced by a continuous function [40].

Transformations on data and equivariance. We assume that data can be transformed through maps from X to X which are Φ-preserving homeomorphisms, namely, homeomorphisms $g : X \to X$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$. The set of these homeomorphisms will be denoted by the symbol $\mathrm{Homeo}_\Phi(X)$.

The following Proposition 2 implies that $\mathrm{Homeo}_\Phi(X)$ is exactly the set of all bijections $g : X \to X$ such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$. Before stating the proposition, it is necessary to observe that an isometry between pseudo-metric spaces can be defined naturally, by generalising the concept of isometry between metric spaces. Let (X_1, d_1) and (X_2, d_2) be two pseudo-metric spaces. It is easy to check that if $f : X_1 \longrightarrow X_2$ is a function verifying the equality $d_1(x, y) = d_2(f(x), f(y))$ for every $x, y \in X_1$, then f is continuous with respect to the topologies induced by d_1 and d_2. If f verifies the previous equality and is bijective, we say that it is an *isometry* between the considered pseudo-metric spaces. If f is an isometry, it is clear that f^{-1} is also an isometry, and that f is a homeomorphism.

Proposition 2. *If g is a bijection from X to X such that $\varphi \circ g \in \Phi$ and $\varphi \circ g^{-1} \in \Phi$ for every $\varphi \in \Phi$, then g is an isometry (and hence a homeomorphism) with respect to D_X.*

Proof. Supplementary Methods, Proposition 2.3. □

We now consider a subgroup G of the group $\mathrm{Homeo}_\Phi(X)$. G represents the set of transformations on data for which we require equivariance to be respected. We can prove (see **Supplementary Methods**, Theorem 2.7) that G is a topological group with respect to the topology induced by the pseudo-metric defined by setting

$$D_G(g_1, g_2) := \sup_{\varphi \in \Phi} D_\Phi(\varphi \circ g_1, \varphi \circ g_2) \tag{3}$$

for every $g_1, g_2 \in G$, and that the action of G on Φ through right composition is continuous. D_G measures the distance between two homeomorphisms as the difference of their actions on the set Φ of possible measurements.

In general G is not compact with respect to the topology induced by D_G, even if Φ is compact. For example, if Φ is the set containing all the 1-Lipschitz functions from $X = S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ to $[0, 1]$, and G is the group of all rotations $\rho_{2\pi q}$ of X of $2\pi q$ radians with q rational number, then the space (G, D_G) is not compact. However, the next result holds.

Theorem 3. *If G is complete then it is also compact with respect to D_G.*

Proof. Supplementary Methods, Theorem 2.8. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

From now on we will suppose that G is complete (and hence compact) with respect to the topology induced by D_G.

The pair (Φ, G) is called a *perception pair*. The group G of transformations can be either learned, or fixed as part of prior knowledge.

The natural pseudo-distance. In addition to the pseudo-metric D_Φ, we define another pseudo-distance d_G on the space Φ [16]. It represents the ground truth in our model. Indeed, it allows for comparison between functions and it vanishes for pairs of functions that are equivalent with respect to the action of our group of homeomorphisms G, which expresses the equivalences between data.

Definition 4. The pseudo-distance $d_G : \Phi \times \Phi \to \mathbb{R}$ is defined by setting

$$d_G(\varphi_1, \varphi_2) = \inf_{g \in G} D_\Phi(\varphi_1, \varphi_2 \circ g). \qquad\qquad (4)$$

It is called the *natural pseudo-distance* associated with the group G acting on Φ.

If $G = \{\mathrm{Id} : x \mapsto x\}$, then d_G equals the sup-norm distance D_Φ on Φ. If G_1 and G_2 are subgroups of $\mathrm{Homeo}_\Phi(X)$ and $G_1 \subseteq G_2$, then the definition of d_G implies that

$$d_{\mathrm{Homeo}_\Phi(X)}(\varphi_1, \varphi_2) \leq d_{G_2}(\varphi_1, \varphi_2) \leq d_{G_1}(\varphi_1, \varphi_2) \leq D_\Phi(\varphi_1, \varphi_2) \qquad (5)$$

for every $\varphi_1, \ \varphi_2 \in \Phi$.

The choice of grounding our approach on homeomorphisms could appear restrictive. However, it is important to notice that the spaces required to be homeomorphic are not the data *per se*, but their support. For instance, grayscale images are represented as functions from $\mathbb{R}^2$ to $\mathbb{R}$, and it is not relevant if the objects represented in them are non-homeomorphic, but simply that the functions are defined on homeomorphic spaces, $\mathbb{R}^2$ in this example.

Group Equivariant Non-Expansive Operators. The operators on data are defined as Group Equivariant Non-Expansive Operators (GENEOs).

Assume that two perception pairs (Φ, G), (Ψ, H) are given together with a fixed homomorphism $T : G \to H$. Each function $F : \Phi \to \Psi$ such that $F(\varphi \circ g) = F(\varphi) \circ T(g)$ for every $\varphi \in \Phi, g \in G$ is said to be a *perception map* from (Φ, G) to (Ψ, H) associated with the homomorphism T. For brevity and when no ambiguity can arise, we will also say that F is a

group equivariant operator. If $(\Phi, G) = (\Psi, H)$ and T is equal to the identity homomorphism $I : G \longrightarrow G$, we can say that F is a G-map.

We observe that the functions in Φ and the functions in Ψ are defined on spaces that are generally different from each other, and the groups of invariance can be different as well. This is important, as it allows one to compose operators hierarchically.

Definition 5. Assume that $(\Phi, G),(\Psi, H)$ are two perception pairs and that a homomorphism $T : G \to H$ has been fixed. If F is a perception map from (Φ, G) to (Ψ, H) with respect to T and F is non-expansive (i.e., $D_\Psi (F(\varphi_1), F(\varphi_2)) \leq D_\Phi (\varphi_1, \varphi_2)$ for every $\varphi_1, \varphi_2 \in \Phi$), then F is called a *Group Equivariant Non-Expansive Operator (GENEO) associated with* $T : G \to H$.

Example 6. As a reference for the reader, we give the following basic example of GENEO. Let Φ be the set containing all 1-Lipschitz functions from $X = S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ to $[0, 1]$, and G be the group of all rotations of S^2 around the z-axis. Let Ψ be the set containing all 1-Lipschitz functions from $Y = S^1 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1 \text{ and } z = 0\}$ to $[0, 1]$, and H be the group of all rotations of S^1. We observe that (Φ, G) and (Ψ, H) are two perception pairs. Now, let us consider the map $F : \Phi \to \Psi$ taking each function $\varphi \in \Phi$ to the function *average on meridians* $\psi \in \Psi$ defined by setting $\psi(\theta) := \frac{1}{\pi} \int_0^\pi \varphi(\theta, \alpha) \, d\alpha$ (where θ and α are the azimuthal and polar angles, respectively), and the homomorphism T taking the rotation of S^2 of θ radians around the z-axis positively oriented to the counter-clock rotation of θ radians of S^1. We can easily check that F is a perception map and a GENEO from (Φ, G) to (Ψ, H), associated with the homomorphism T. In this example F and T are surjective, but an example where F and T are not surjective can be easily found, e.g. by restricting Φ to the singleton $\bar{\Phi}$ containing only the null function and G to the trivial group $\bar{G}$ containing only the identical homomorphism.

Compactness and convexity of the space of GENEOs. We can prove that if the function spaces are compact and convex, then the space of GENEOs is compact and convex too. The compactness guarantees that the space of GENEOs can be approximated by a finite set. The convexity implies that new GENEOs can be obtained by convex combinations of pre-existing GENEOs.

If $\mathcal{F}^{\mathrm{all}} := \mathrm{GENEO}\,((\Phi, G), (\Psi, H))$ denotes the set of GENEOs between two perception pairs (Φ, G), (Ψ, H) associated with the homomorphism $T : G \to H$, then the following theorem holds:

Theorem 7. *If Φ and Ψ are compact with respect to D_Φ and D_Ψ, respectively, then $\mathcal{F}^{\mathrm{all}}$ is compact with respect to the pseudo-metric*

$$D_{\mathrm{GENEO}}\,(F_1, F_2) := \sup_{\varphi \in \Phi} D_\Psi \,(F_1(\varphi), F_2(\varphi)) . \tag{6}$$

Proof. Supplementary Methods, Theorem 3.1. $\qquad\qquad\square$

Now let $F_1, F_2, \dots, F_n \in \mathcal{F}^{\mathrm{all}}$. If $(a_1, a_2, \dots, a_n) \in \mathbb{R}^n$ with $\sum_{i=1}^n |a_i| \leq 1$ and Ψ is convex, then

$$F_\Sigma(\varphi) := \sum_{i=1}^n a_i F_i(\varphi) \tag{7}$$

is a GENEO from (Φ, G) to (Ψ, H) with respect to T. Therefore, the following theorem holds:

Theorem 8. *If Ψ is convex, then the set of GENEOs from (Φ, G) to (Ψ, H) with respect to T is convex.*

Proof. Supplementary Methods, Theorem 3.3.   □

Pseudo-metrics induced by persistent homology. To compare data under the action of a set $\mathcal{F}$ of GENEOs, one could simply define, for $\varphi_1, \varphi_2 \in \Phi$,

$$D_{\mathcal{F},\Phi}(\varphi_1, \varphi_2) := \sup_{F \in \mathcal{F}} \|F(\varphi_1) - F(\varphi_2)\|_\infty. \tag{8}$$

Though, the computation of $D_{\mathcal{F},\Phi}(\varphi_1, \varphi_2)$ for every pair (φ_1, φ_2) of admissible functions is computationally expensive. Persistent homology allows us to replace $D_{\mathcal{F},\Phi}$ with a pseudo-metric $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$ computationally more efficient, but still stable and strongly invariant. Where, a pseudo-metric $\hat{d}$ on Φ is *strongly G-invariant* if it is invariant under the action of G with respect to each variable, that is, if

$$\hat{d}(\varphi_1, \varphi_2) = \hat{d}(\varphi_1 \circ g, \varphi_2) = \hat{d}(\varphi_1, \varphi_2 \circ g) = \hat{d}(\varphi_1 \circ g, \varphi_2 \circ g)$$

for every $\varphi_1, \varphi_2 \in \Phi$ and every $g \in G$.

We briefly recall that persistent homology allows to represent the topological and geometrical features of a topological space X (e.g. an image, a 3-dimensional mesh) as it is *seen* by a continuous, real-valued function φ defined on the space. The homology functor (see for instance [41]) is used to encode the information of the pair (X, φ) in the form of *persistence diagrams*. In other words, for every $k \in \mathbb{N}$ we can associate each continuous function $\varphi : X \to \mathbb{R}$ with a persistence diagram $\mathbb{D}_\varphi^k$, that is represented by a discrete collection of points in the real plane. See Figure 2. Persistence diagrams can be represented by suitable functions called persistent Betti numbers and denoted by $r_k(\varphi)$. The matching distance δ_{match} between persistence diagrams depicted in panel C of Figure 2, corresponds to the metric d_{match} between persistent Betti numbers. For details see [42]. Beyond the technicalities that are needed to define the concept of persistence diagram, two important points are to be stressed. First, persistent Betti numbers can be rapidly computed. Second, the distance d_{match} between persistent Betti numbers is efficiently computable and gives a lower bound for the max-norm distance between functions: $\delta_{\text{match}}(\mathbb{D}_{\varphi_1}^k, \mathbb{D}_{\varphi_2}^k) = d_{\text{match}}(r_k(\varphi_1), r_k(\varphi_2)) \leq \|\varphi_1 - \varphi_2\|_\infty$. It follows that the distance d_{match} can be used as an efficient proxy for the max-norm distance between real-valued functions. We refer the reader to [43, 44, 45] for further details.

Let us consider a subset $\mathcal{F} \neq \emptyset$ of $\mathcal{F}^{\text{all}}$. For every fixed k, we define the pseudo-metric $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$ on Φ:

$$\mathcal{D}_{\text{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2) := \sup_{F \in \mathcal{F}} d_{\text{match}}(r_k(F(\varphi_1)), r_k(F(\varphi_2))) \tag{9}$$

for every $\varphi_1, \varphi_2 \in \Phi$, where $r_k(\varphi)$ denotes the kth persistent Betti numbers function with respect to the function $\varphi : X \to \mathbb{R}$. Observe that this pseudo-metric is an optimal lower bound for the metric defined in Equation (8). Then:

Proposition 9. $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ *is a strongly G-invariant pseudo-metric on Φ.*

Proof. Supplementary Methods, Proposition 4.3. $\qquad\square$

The pseudo-distance $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ is stable with respect to both the pseudo-metric d_G and the metric D_Φ. This fact guarantees that $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ can be used in the presence of noise as it is stated in the following result.

Theorem 10. *If $\mathcal{F}$ is a non-empty subset of $\mathcal{F}^{\mathrm{all}}$, then*

$$\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}} \le d_G \le D_\Phi. \tag{10}$$

Proof. Supplementary Methods, Theorem 4.5. $\qquad\square$

The definitions of the natural pseudo-distance d_G and the pseudo-distance $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ come from different theoretical concepts. The former is based on a variational approach involving the set of all homeomorphisms in G, while the latter refers only to a comparison of persistent homologies depending on a family of group equivariant non-expansive operators. Thus, the next result may appear unexpected. Indeed, the role of the group G is not explicitly expressed in the definition of $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$, but implicitly encoded in the GENEOs that are equivariant with respect to G. Moreover, the information contained in each single persistence diagram used in the definition of $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ is generally much smaller than the one expressed by the natural pseudo-distance d_G.

Theorem 11. *Let us assume that $\Phi = \Psi$, every function in Φ is non-negative, the k-th Betti number of X does not vanish, and Φ contains each constant function c for which a function $\varphi \in \Phi$ exists such that $0 \le c \le \|\varphi\|_\infty$. Then $\mathcal{D}^{\mathcal{F}^{\mathrm{all}},k}_{\mathrm{match}} = d_G$.*

Proof. Supplementary Methods, Theorem 4.6. $\qquad\square$

We observe that if Φ is bounded, the assumption that every function in Φ is non-negative is not quite restrictive. Indeed, we can obtain it by adding a suitable constant value to every admissible function.

Here we show how $\mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}$ can be approximated arbitrarily well with a finite subset of operators:

Proposition 12. *Let $\mathcal{F}$ be a non-empty subset of $\mathcal{F}^{\mathrm{all}}$. For every $\varepsilon > 0$, a finite subset $\mathcal{F}^*$ of $\mathcal{F}$ exists, such that*

$$|\mathcal{D}^{\mathcal{F}^*,k}_{\mathrm{match}}(\varphi_1, \varphi_2) - \mathcal{D}^{\mathcal{F},k}_{\mathrm{match}}(\varphi_1, \varphi_2)| \le \varepsilon \tag{11}$$

for every $\varphi_1, \varphi_2 \in \Phi$.

Proof. Supplementary Methods, Proposition 4.8. $\qquad\square$

Since the compactness of the space $\mathcal{F}^{\text{all}}$ guarantees we can cover $\mathcal{F}$ by a finite set of balls in $\mathcal{F}^{\text{all}}$ of radius ε, centered at points of a finite set $\mathcal{F}' \subseteq \mathcal{F}$, the proposition above states that the approximation of $\mathcal{D}_{\text{match}}^{\mathcal{F},k}(\varphi_1, \varphi_2)$ can be reduced to the computation of $\mathcal{D}_{\text{match}}^{\mathcal{F}',k}(\varphi_1, \varphi_2)$, i.e. the maximum of a finite set of bottleneck distances between persistence diagrams, which are well-known to be computable by means of efficient algorithms.

Finally, we use persistent homology to define a computable and stable pseudo-metric between GENEOs. If $F_1, F_2 \in \mathcal{F}^{\text{all}}$, for every fixed $k \in \mathbb{N}$, we can set

$$\Delta_{\text{GENEO}}(F_1, F_2) := \sup_{\varphi \in \Phi} d_{\text{match}}(r_k(F_1(\varphi)), r_k(F_2(\varphi))) \tag{12}$$

for which it holds $\Delta_{\text{GENEO}} \leq D_{\text{GENEO}}$. The approximations of the pseudo-metrics Δ_{GENEO} and $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$ are used in the experiments shown.

As a final remark, we observe that our approach based on GENEOs and persistent homology can be used also when we wish to have equivariance with respect to a *set* instead of a *group* of homeomorphisms. Indeed, whereas the definition of the natural pseudo-distance d_G requires that G has the structure of a group, the definition of $\mathcal{D}_{\text{match}}^{\mathcal{F},k}$ does not need this assumption.

Experimental setting

Isometry equivariant non-expansive operators. We define a parametric family of non-expansive operators which are equivariant with respect to Euclidean plane isometries.

Given $\sigma > 0$ and $\tau \in \mathbb{R}$, we consider the 1-dimensional Gaussian function with width σ and centre τ

$$g_\tau(t) := e^{-\frac{(t-\tau)^2}{2\sigma^2}},$$

where $g_\tau : \mathbb{R} \to \mathbb{R}$. For a positive integer k, we take the set S of the $2k$-tuples $(a_1, \tau_1, \ldots, a_k, \tau_k) \in \mathbb{R}^{2k}$ for which $\sum_{i=1}^{k} a_i^2 = \sum_{i=1}^{k} \tau_i^2 = 1$. S is a submanifold of $\mathbb{R}^{2k}$.

For each $p = (a_1, \tau_1, \ldots, a_k, \tau_k) \in S$, we then consider the function $G_p : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$G_p(x, y) := \sum_{i=1}^{k} a_i g_{\tau_i}\left(\sqrt{x^2 + y^2}\right).$$

We denote by F_p the convolutional operator mapping each continuous function with compact support $\varphi : \mathbb{R}^2 \to \mathbb{R}$ to the continuous and compactly supported function $\psi : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\psi(x, y) := \int_{\mathbb{R}^2} \varphi(\alpha, \beta) \cdot \frac{G_p(x - \alpha, y - \beta)}{\|G_p\|_{L^1}} \, d\alpha \, d\beta.$$

Then, the operator F_p is a group equivariant non-expansive operator with respect to the group I of Euclidean plane isometries. We call F_p a IENEO (Isometry Equivariant Non-Expansive Operator).

The IENEO F_p is parametric with respect to the $2k$-tuple $p = (a_1, \tau_1, \ldots, a_k, \tau_k) \in S$. Therefore, we define a parametric family of IENEOs $\mathcal{F} = \{F_p\}_{p \in S}$.

Image preprocessing. Images are preprocessed according to the pipeline described in the first column of Figure 3. Every image ι is first reshaped to size $(128, 128)$, then blurred with a 3×3 Gaussian kernel and finally standardised as $\iota_s = \frac{\iota - \text{mean}(\iota)}{\sigma(\iota)}$. The same preprocessing is applied in all experiments and to all datasets.

Operators selection and sampling on labelled datasets. Let $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ be a dataset equipped with a labelling function $l : \Phi \to J \subset \mathbb{N}$. We assume that the dataset can be written as the disjoint union $\Phi = \sqcup_{i \in J} \Phi_i$ where Φ_i contains samples labelled by i. Let $\mathcal{F}$ be the space of operators that will act on the samples. We begin by randomly sampling N candidate operators in $\mathcal{F}$, let us denote them as the set $\mathcal{C} = \{F_k\}_{k \in \{1, \ldots, N\}}$. We then select those operators that consider as similar the objects belonging to the same class. Let us consider the samples in Φ_l, for each of the candidate operators $F \in \mathcal{C}$, we define the label-dependent value

$$s_l(F) = \max_{\varphi_i^l, \varphi_j^l} d_{\text{match}}\left(r_k\left(F\left(\varphi_i^l\right)\right), r_k\left(F\left(\varphi_j^l\right)\right)\right),$$

in these experiments we set $k = 1$, although any degree could be used. A candidate operator F is *selected* if $s_l(F)$ is smaller than a fixed threshold ϵ for every l. Let us denote by $\mathcal{S}$ the set of selected operators.

As the selection criterion does not guarantee that the operators are maximally diverse, when evaluated within and in-between classes, we sample the elements of $\mathcal{S}$ which focus on different shape characteristics. Given a class l, we define the distance between two operators F_p and F_q (cf. Equation (12))

$$\Delta_{\text{GENEO}}^l(F_p, F_q) := \max_{\varphi_i^l} d_{\text{match}}\left(r_k\left(F_p\left(\varphi_i^l\right)\right), r_k\left(F_q\left(\varphi_i^l\right)\right)\right).$$

For every label l, we sort the pairs (F_p, F_q) in ascending order of Δ_{GENEO}^l, and assign to each pair of operators its index in the sorted list of distances. We then define the *interclass contrastive score* of the pair (F_p, F_q) as the sum of its indices over all classes. Finally, we remove from $\mathcal{S}$ redundant operators, i.e. we select only one operators for pairs whose score is below a fixed threshold t.

Finally, two objects φ_1 and φ_2 can be compared by computing the strongly G-invariant pseudo-metric $\mathcal{D}_{\text{match}}^{\mathcal{S},1}(\varphi_1, \varphi_2)$, defined in Equation (9).

Metric learning through selection and sampling. Selected and sampled operators $F_i \in \mathcal{S}$ can be used to measure distances between pairs of validation samples as

$$d_{\mathcal{S}}(\varphi, \varphi') = \max_{F \in \mathcal{S}} d_{\text{match}}(r_1(F(\varphi)), r_1(F(\varphi'))). \tag{13}$$

Note that $d_{\mathcal{S}}$ is invariant with respect to the action of the group of planar isometries. This invariance is naturally inherited by the usage of d_{match}.

Competing interests

References

[1] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al., Mastering the game of go without human knowledge, Nature 550 (7676) (2017) 354.

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115.

[3] Y. LeCun, Y. Bengio, et al., Convolutional networks for images, speech, and time series, The handbook of brain theory and neural networks 3361 (10) (1995) 1995.

[4] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, Neural computation 29 (9) (2017) 2352–2449.

[5] Y. Kim, Convolutional neural networks for sentence classification, arXiv preprint arXiv:1408.5882.

[6] P. Niyogi, Manifold regularization and semi-supervised learning: Some theoretical analyses, The Journal of Machine Learning Research 14 (1) (2013) 1229–1250.

[7] F. Anselmi, L. Rosasco, T. Poggio, On invariance and selectivity in representation learning, Information and Inference: A Journal of the IMA 5 (2) (2016) 134–158. `arXiv:http://oup.prod.sis.lan/imaiai/article-pdf/5/2/134/6990886/iaw009.pdf`, `doi:10.1093/imaiai/iaw009`.
URL `https://doi.org/10.1093/imaiai/iaw009`

[8] T. Cohen, M. Welling, Group equivariant convolutional networks, in: International conference on machine learning, 2016, pp. 2990–2999.

[9] D. E. Worrall, S. J. Garbin, D. Turmukhambetov, G. J. Brostow, Harmonic networks: Deep translation and rotation equivariance, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2017.

[10] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828. `doi:10.1109/TPAMI.2013.50`.
URL `http://dx.doi.org/10.1109/TPAMI.2013.50`

[11] N. Tishby, N. Zaslavsky, Deep learning and the information bottleneck principle, in: 2015 IEEE Information Theory Workshop (ITW), 2015, pp. 1–5. `doi:10.1109/ITW.2015.7133169`.

[12] S. Mallat, Group invariant scattering, Communications on Pure and Applied Mathematics 65 (10) (2012) 1331–1398. `arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.21413`, `doi:10.1002/cpa.21413`.
URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21413`

[13] S. Mallat, Understanding deep convolutional networks, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374 (2065) (2016) 20150203. `arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2015.0203`, `doi:10.1098/rsta.2015.0203`.
URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2015.0203`

[14] C. Zhang, S. Voinea, G. Evangelopoulos, L. Rosasco, T. Poggio, Discriminative template learning in group-convolutional networks for invariant speech representations, Vol. 2015-January, 2015, pp. 3229–3233.
URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959146672&partnerID=40&md5=97cf2163120c30c46e0c90e5472223fb`

[15] F. Anselmi, G. Evangelopoulos, L. Rosasco, T. Poggio, Symmetry-adapted representation learning, Pattern Recognition 86 (2019) 201 – 208. `doi:https://doi.org/10.1016/j.patcog.2018.07.025`.
URL `http://www.sciencedirect.com/science/article/pii/S0031320318302620`

[16] P. Frosini, G. Jabłoński, Combining persistent homology and invariance groups for shape comparison, Discrete Comput. Geom. 55 (2) (2016) 373–409. `doi:10.1007/s00454-016-9761-y`.
URL `http://dx.doi.org/10.1007/s00454-016-9761-y`

[17] P. Frosini, N. Quercioli, Some remarks on the algebraic properties of group invariant operators in persistent homology, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10410 LNCS (2017) 14–24. `doi:10.1007/978-3-319-66808-6_2`.
URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85029005197&doi=10.1007%2f978-3-319-66808-6_2&partnerID=40&md5=d59c082eca0f7ca45fbb0e4ac52336b2`

[18] F. Camporesi, P. Frosini, N. Quercioli, On a new method to build group equivariant operators by means of permutants, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11015 LNCS (2018) 265–272. `doi:10.1007/978-3-319-99740-7_18`.
URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053191136&doi=10.1007%2f978-3-319-99740-7_18&partnerID=40&md5=2eab1ec2cc39fc819262e1572d25ce65`

[19] S. Smale, L. Rosasco, J. Bouvrie, A. Caponnetto, T. Poggio, Mathematics of the neural response, Foundations of Computational Mathematics 10 (1) (2010) 67–91. `doi:10.1007/s10208-009-9049-1`.
URL `https://doi.org/10.1007/s10208-009-9049-1`

[20] G. Carlsson, Topology and data, Bull. Amer. Math. Soc. (N.S.) 46 (2) (2009) 255–308. `doi:10.1090/S0273-0979-09-01249-X`.
URL `https://doi.org/10.1090/S0273-0979-09-01249-X`

[21] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, G. E. Carlsson, Extracting insights from the shape of complex data using topology, in: Scientific reports, Vol. 3, 2013.

[22] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, L. Ziegelmeier, Persistence images: A stable vector representation of persistent homology, J. Mach. Learn. Res. 18 (1) (2017) 218–252.
URL `http://dl.acm.org/citation.cfm?id=3122009.3122017`

[23] C. S. Pun, K. Xia, S. Xian Lee, Persistent-homology-based machine learning and its applications – A survey, arXiv e-prints (2018) arXiv:1811.00252`arXiv:1811.00252`.

[24] R. B. Gabrielsson, G. Carlsson, Exposition and interpretation of the topology of neural networks, CoRR abs/1810.03234. `arXiv:1810.03234`.
URL `http://arxiv.org/abs/1810.03234`

[25] G. Carlsson, R. B. Gabrielsson, Topological approaches to deep learning (2018). `arXiv:1811.01122`.

[26] A. Zomorodian, Fast construction of the Vietoris-Rips complex, Computers & Graphics 34 (3) (2010) 263–271.

[27] R. Fabbri, L. D. F. Costa, J. C. Torelli, O. M. Bruno, 2D Euclidean distance transform algorithms: A comparative survey, ACM Computing Surveys (CSUR) 40 (1) (2008) 2.

[28] L. Deng, The mnist database of handwritten digit images for machine learning research [best of the web], IEEE Signal Processing Magazine 29 (6) (2012) 141–142.

[29] H. Xiao, K. Rasul, R. Vollgraf, Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, arXiv preprint arXiv:1708.07747.

[30] A. Krizhevsky, V. Nair, G. Hinton, The cifar-10 dataset, online: http://www. cs. toronto. edu/kriz/cifar. html 55.

[31] A. Krizhevsky, G. Hinton, et al., Learning multiple layers of features from tiny images, Tech. rep., Citeseer (2009).

[32] P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R, Bioinformatics 24 (5) (2007) 719–720.

[33] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010, pp. 249–256.

[34] G. E. Hinton, A. Krizhevsky, S. D. Wang, Transforming auto-encoders, in: International Conference on Artificial Neural Networks, Springer, 2011, pp. 44–51.

[35] S. Sabour, N. Frosst, G. E. Hinton, Dynamic Routing Between Capsules, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 3856–3866.
URL `http://papers.nips.cc/paper/6975-dynamic-routing-between-capsules.pdf`

[36] P. Frosini, Towards an Observer-oriented Theory of Shape Comparison, in: A. Ferreira, A. Giachetti, D. Giorgi (Eds.), Eurographics Workshop on 3D Object Retrieval, The Eurographics Association, 2016. `doi:10.2312/3dor.20161080`.

[37] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, T. R. Evans, Reconstruction and representation of 3d objects with radial basis functions, in: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '01, ACM, New York, NY, USA, 2001, pp. 67–76. `doi:10.1145/383259.383266`.
URL `http://doi.acm.org/10.1145/383259.383266`

[38] M. Atzmon, H. Maron, Y. Lipman, Point convolutional neural networks by extension operators, ACM Trans. Graph. 37 (4) (2018) 71:1–71:12. `doi:10.1145/3197517.3201301`.
URL `http://doi.acm.org/10.1145/3197517.3201301`

[39] A. Collins, A. Zomorodian, G. Carlsson, L. Guibas, A Barcode Shape Descriptor for Curve Point Cloud Data, in: M. Gross, H. Pfister, M. Alexa, S. Rusinkiewicz (Eds.), SPBG'04 Symposium on Point - Based Graphics 2004, The Eurographics Association, 2004. `doi:10.2312/SPBG/SPBG04/181-191`.

[40] B. Di Fabio, P. Frosini, Filtrations induced by continuous functions, Topology and its Applications 160 (12) (2013) 1413–1422.

[41] A. Hatcher, Algebraic topology, Cambridge University Press, Cambridge, 2002.

[42] A. Cerri, B. D. Fabio, M. Ferri, P. Frosini, C. Landi, Betti numbers in multidimensional persistent homology are stable functions, Mathematical Methods in the Applied Sciences 36 (12) (2013) 1543–1557.

[43] S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, M. Spagnuolo, Describing Shapes by Geometrical-topological Properties of Real Functions, ACM Comput. Surv. 40 (4) (2008) 12:1–12:87. `doi:10.1145/1391729.1391731`.
URL `http://doi.acm.org/10.1145/1391729.1391731`

[44] G. Carlsson, A. Zomorodian, The theory of multidimensional persistence, Discrete Comput. Geom. 42 (1) (2009) 71–93. `doi:10.1007/s00454-009-9176-0`.
URL `http://dx.doi.org/10.1007/s00454-009-9176-0`

[45] H. Edelsbrunner, J. Harer, Persistent homology—a survey, in: Surveys on discrete and computational geometry, Vol. 453 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2008, pp. 257–282. `doi:10.1090/conm/453/08802`.
URL `http://dx.doi.org/10.1090/conm/453/08802`